USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA USC

Aida
Carballo Fazanes

Tese de doutoramento

Evaluación fiable de las habilidades motrices básicas en la infancia. Proyecto ALFA-MOV

Santiago de Compostela, 2022

Programa de Doutoramento en Investigación Clínica en Medicina

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

TESE DE DOUTORAMENTO

EVALUACIÓN FIABLE DE LAS HABILIDADES MOTRICES BÁSICAS EN LA INFANCIA. PROYECTO ALFA-MOV

Aida Carballo Fazanes

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN INVESTIGACIÓN CLÍNICA EN MEDICINA

SANTIAGO DE COMPOSTELA
2022

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DA AUTORA DA TESE

EVALUACIÓN FIABLE DE LAS HABILIDADES MOTRICES BÁSICAS EN LA INFANCIA. PROYECTO ALFA-MOV

Dna. Aida Carballo Fazanes

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) *A tese abarca os resultados da elaboración do meu traballo.*
2) *De ser o caso, na tese faise referencia ás colaboración que tivo este traballo.*
3) *Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.*
4) *A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico.*

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En *Santiago de Compostela*, a *04 de outubro de 2022*.

Asdo. Aida Carballo Fazanes

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR / TITOR DA TESE

EVALUACIÓN FIABLE DE LAS HABILIDADES MOTRICES BÁSICAS EN LA INFANCIA. PROYECTO ALFA-MOV

D. Antonio Rodríguez Núñez

D. Cristian Abelairas Gómez

INFORMAN:

Que a presente tese, correspóndese co traballo realizado por Dna. Aida Carballo Fazanes, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de Monográfica con reprodución de publicacións, nos que a participación da doutoranda foi decisiva para a súa elaboración e as publicacións axústanse ao Plan de Investigación.

En *Santiago de Compostela*, a *04 de outubro de 2022*.

Asdo. Antonio Rodríguez Núñez
(Titor e director)

Asdo. Cristian Abelairas Gómez
(Director)

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CONFLITOS DE INTERESE

EVALUACIÓN FIABLE DE LAS HABILIDADES MOTRICES BÁSICAS EN LA INFANCIA. PROYECTO ALFA-MOV

Este traballo forma parte do proxecto de investigación "*Alfabetización Motriz y estilos de vida saludables: Desarrollo de una herramienta informática de evaluación de la competencia motriz de niños y niñas en Educación Primaria*" (RTI2018-096106-A-I00) financiado polo MCIN/AEI /10.13039/501100011033/ e por FEDER "Una manera de hacer Europa".

Proxecto *RTI2018-096106-A-I00* financiado por:



Ademais, a doutoranda recibiu financiación do Ministerio de Universidades a través do programa "*Ayudas para la formación del profesorado universitario (FPU), en el marco del Plan Estatal de Investigación Científica y Técnica de Innovación 2017-2020*" (FPU19/02917).

En *Santiago de Compostela*, a *05 de outubro de 2022*.

Asdo. Aida Carballo Fazanes

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

PRINCIPIOS ÉTICOS

EVALUACIÓN FIABLE DE LAS HABILIDADES MOTRICES BÁSICAS EN LA INFANCIA. PROYECTO ALFA-MOV

Dna. Aida Carballo Fazanes

Os estudos presentados neste traballo leváronse a cabo de acordo aos principios éticos recollidos na Declaración de Helsinki, e declaro que:

1) *A participación foi voluntaria e obtívose o consentimento informado dos pais/nais/titores legais de todos os nenos participantes.*
2) *Os datos expostos baséanse en información anónima dos participantes.*
3) *O protocolo do estudo foi aprobado polo comité de ética da facultade de Ciencias da Educación e do Deporte da Universidade de Vigo (Código: 17-0320).*

En Santiago de Compostela, a *04 de outubro de 2022*.

Asdo. Aida Carballo Fazanes

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A meus pais,
polo que son.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A totes les pors,
als petits errors
que han construït
el que som ara aquí.
Nous camins vindran,
I amb ells vindran nous somnis.

Txarango

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AGRADECEMENTOS

Aos meus directores, Cristian e Antonio, por ser o motor desta viaxe. Con vós comezou todo. Grazas por confiar sempre no meu traballo, por introducirme e acompañarme no mundo da investigación e docencia, por traballar de xeito incansable. Grazas por facerme medrar tanto profesional como persoalmente.

A meus pais, Chus e Camilo, o meu maior apoio. Grazas por ser o inesgotable combustible que me permite voar alto e manterme sempre arriba. Sen vós, isto non sería posible.

Da igual o destino que elixa, o importante son os compañeiros e compañeiras de viaxe. A Cristina, Geno, Zequi e Santi, por ser a miña guía e por acompañarme nos momentos de maior desafío.

Grazas a todo o equipo de ALFA-MOV e a todas as persoas integrantes de CLINURSID e SICRUS por compartir comigo os anos máis enriquecedores da miña vida. En especial, a miña familia valenciana, AFIPS; e da ESDL (Portugal).

A Carmen, Bárbara, Mateo, Alba, Carla e Ale. Por estar sempre dispostas a brindar ou abrazar (dependendo do momento), e facendo esta etapa, e, en xeral, a miña vida, máis fácil e moito mellor.

A Borja, Bea, Óscar e María, os protagonistas dos mellores recordos da miña vida. Grazas por agarrarme da man e ensinarme que se chega lonxe camiñando despacio, e si, sempre que sexa con vós tirando forte.

A Alberto, por ser a peza clave durante os anos máis duros. Polo sorriso constante, paciencia infinita e amor incondicional.

Ás mulleres da miña vida: Lidia, Ana, Mari, Mucha, Laira e María. Por acompañarme e impulsarme no camiño desde antes de que tivera memoria, sempre orgullosas dos meus pasos. En especial, a Alexandra e Antía por ser revolución e a miña inspiración.

A Chelo e Manolo, meus avós. Exemplos a seguir, a miña luz, guía e vida.

En especial, ao avó Camilo e a Nandi, por facerme brillar desde o máis alto.

En definitiva, grazas a todas as persoas que me acompañastes no camiño e fixestes que esta etapa merecese a pena. Chegamos xuntas pero, en realidade, o percorrido só acaba de comezar.

Pels infinits camins del món ens seguirem trobant. [Txarango]

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ÍNDICE

ABREVIATURAS **21**

ÍNDICE DE TABLAS **23**

ÍNDICE DE FIGURAS **25**

1 RESUMEN, RESUMO Y ABSTRACT **27**

1.1 RESUMEN 27

1.2 RESUMO 29

1.3 ABSTRACT 31

2 INTRODUCCIÓN **33**

2.1 ESTILOS DE VIDA SALUDABLES 33

2.1.1 Inactividad física, sedentarismo y obesidad: el problema *33*

2.1.2 La actividad física como agente clave en la adquisición de un estilo de vida saludable *35*

2.2 ALFABETIZACIÓN MOTRIZ 37

2.2.1 Definición *37*

2.2.2 Evolución de los modelos conceptuales de desarrollo *38*

2.2.3 Dominios de la alfabetización motriz *44*

2.2.3.1 Dominio Físico 45

2.2.3.2 Dominio Psicológico 45

2.2.3.3 Dominio Cognitivo 45

2.2.3.4 Dominio Social 46

2.3 COMPETENCIA MOTRIZ 46

2.4 HABILIDADES MOTRICES BÁSICAS 47

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.4.1 Definición *47*
2.4.2 Clasificación *48*
2.5 EVALUACIÓN DE LAS HABILIDADES MOTRICES BÁSICAS 49
2.5.1 Importancia del conocimiento y seguimiento del grado de desarrollo de las HMB *50*
2.5.2 La evaluación sistemática de las habilidades motrices básicas: una carencia a solventar *51*
2.5.3 Test of Gross Motor Development (TGMD) *54*
2.5.4 La fiabilidad como requisito imprescindible en la evaluación de las HMB *57*
2.6 JUSTIFICACIÓN 58

3 HIPÓTESIS 59

4 OBJETIVOS 61

5 METODOLOGÍA 63

6 RESULTADOS 65
6.1 RESULTADOS DERIVADOS DE LA TESIS DOCTORAL 65
6.1.1 Artículo 1. Fiabilidad del Test of Gross Motor Development: una revisión sistemática *65*
6.1.2 Artículo 2. Fiabilidad intra-observador (evaluación en directo vs vídeo) e inter-observador (evaluadores expertos vs novatos) del Test of Gross Motor Development – Third Edition *112*
6.1.3 Artículo 3: Fiabilidad inter-observador del Test of Gross Motor Development – Third Edition, tras el consenso previo de los evaluadores en los criterios de rendimiento *140*
6.2 *ALFAMOV*: LA HERRAMIENTA PARA LA EVALUACIÓN DE LAS HABILIDADES MOTRICES BÁSICAS 164

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.3 PUBLICACIONES DIRECTAMENTE RELACIONADAS CON LA TEMÁTICA DE LA TESIS DOCTORAL 173
6.4 OTRAS ACTIVIDADES QUE HAN CONTRIBUIDO A LA FORMACIÓN Y APRENDIZAJE INVESTIGADORA DE LA CANDIDATA 174
6.4.1 Estancias de investigación durante el período predoctoral 174
6.4.2 Otras publicaciones de la candidata durante el período predoctoral (indexadas en JCR) 175
6.4.3 Presentaciones en congresos científicos 177
6.4.4 Premios recibidos: 179

7 DISCUSIÓN 181
7.1 FIABILIDAD DEL TGMD 182
7.1.1 Fiabilidad test-retest 183
7.1.2 Fiabilidad intra-observador 184
7.1.3 Fiabilidad inter-observador 184
7.1.4 Fiabilidad por habilidad 186
7.2 EXPERIENCIA DE LOS EVALUADORES 188
7.3 EVALUACIÓN EN VÍDEO Y EN DIRECTO 190
7.4 CONSENSO ENTRE EVALUADORES 192
7.5 ANÁLISIS ESTADÍSTICO DE FIABILIDAD 194

8 LIMITACIONES Y FORTALEZAS 197
8.1 LIMITACIONES 197
8.2 FORTALEZAS 198

9 IMPLICACIONES PRÁCTICAS 199

10 CONCLUSIONES 201

11 CONCLUSIONS 205

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

12 LÍNEAS FUTURAS DE INVESTIGACIÓN 207

13 REFERENCIAS BIBLIOGRÁFICAS..... 209

ANEXO 1. PUBLICACIONES..... 231

ANEXO 2. ESTANCIAS 275

ANEXO 3. COMITÉ DE ÉTICA 281

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ABREVIATURAS

AF	Actividad física
AM	Alfabetización Motriz
CCI	Coeficiente de correlación intraclase
CM	Competencia motriz
CMG	Coeficiente motor general
COSMIN	COnsensus-based Standards for the selection of health status Measurement INstruments
DT	Desviación Típica
EE	Evaluador experto
EN	Evaluador novato
HMB	Habilidades Motrices Básicas
IVC	Índice de validez de contenido
JCR	Journal Citation Reports
OMS	Organización Mundial de la Salud
P. ej.	Por ejemplo
TEA	Trastorno del Espectro Autista
TGMD	Test of Gross Motor Development

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ÍNDICE DE TABLAS

Tabla 1. Habilidades motrices básicas más comunes. 49

Tabla 2. Principales test de evaluación de las habilidades motrices básicas. 53

Tabla 3. Evolución de las habilidades locomotrices y de control de objetos en las versiones del Test of Gross Motor Development. 55

Tabla 4. Habilidades, criterios de rendimiento y puntuaciones del TGMD-3. 56

Tabla 5. Habilidades motrices básicas incluidas en *AlfaMov*. 169

Tabla 6. Habilidades, criterios de rendimiento y puntuaciones de *AlfaMov*. 170

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ÍNDICE DE FIGURAS

Figura 1. Modelo de desarrollo propuesto por Stodden et al. 39

Figura 2. Modelo conceptual propuesto por Robinson et al.....40

Figura 3 . Modelo de desarrollo de habilidades motrices básicas para la práctica de actividad física a lo largo de la vida propuesto por Hulteen et al.....41

Figura 4 . Modelo conceptual propuesto por Cairney et al. 42

Figura 5 . Modelo propuesto por Barnett et al.. 43

Figura 6. Modelo australiano de AM. 44

Figura 7. Esquema a mano alzada del diseño de navegación. 165

Figura 8. Marca corporativa de *Alfamov* 166

Figura 9. Tipografía seleccionada para *Alfamov*. 166

Figura 10. Gama de colores utilizados en *Alfamov*. 166

Figura 11. Diseño de los iconos de *Alfamov*. 166

Figura 12. Evolución del diseño visual: versión inicial (a), versión intermedia (b), y versión final (c). 167

Figura 13. Comunicación para aplicar cambios entre el equipo de investigación de *Alfamov* y la empresa desarrolladora de software 168

Figura 14. Ejemplos de pantallas del diseño final de *Alfamov*.....171

Figura 15. Ejemplo de perfil de un niño tras evaluar su competencia motriz con *Alfamov*.....172

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1 RESUMEN, RESUMO Y ABSTRACT

1.1 RESUMEN

El incremento de comportamientos sedentarios de nuestra sociedad, así como las tasas elevadas alcanzadas de obesidad y su tendencia al aumento en las últimas décadas, provoca que se continúen buscando y entendiendo sus causas. En este sentido, se han reportado pruebas sobre el papel de la competencia de las habilidades motrices básicas en la práctica de actividad física y, por consiguiente, en la adquisición de estilos de vida saludables a lo largo de la vida. Por ello, es preocupante el descenso que han sufrido los niveles de competencia de mencionadas habilidades en las últimas décadas. Las medidas de intervención podrían enfocarse en tratar de incrementar la competencia motriz de los más jóvenes para conseguir un estilo de vida activo desde la infancia. Para poder implementar esta intervención es necesario conocer el nivel de desarrollo de competencia motriz de los niños y niñas a través de su evaluación, para lo cual se han descrito diferentes barreras. Entre ellas, el tiempo necesario para llevarla a cabo junto a la escasez del mismo para el profesorado y el personal sanitario de atención primaria, la falta de formación o conocimientos entre los principales agentes implicados o la necesidad de herramientas adecuadas para llevar a cabo la evaluación.

En esta tesis doctoral se presentan, entre otros resultados adicionales, dos artículos científicos publicados y uno en proceso de revisión que tratan de dar respuesta a parte de los problemas mencionados. Concretamente, a la fiabilidad de la evaluación de la competencia motriz, requisito imprescindible para realizar una evaluación rigurosa.

La primera publicación es una revisión sistemática sobre la fiabilidad de uno de los instrumentos más utilizados para la evaluación de la competencia motriz gruesa, el *Test of Gross Motor Development*,

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mostrando que es una herramienta psicométrica robusta de evaluación. Los dos estudios restantes (uno de ellos publicado), muestran los resultados de fiabilidad entre observadores de distintas características (novatos y expertos) con dos métodos de evaluación (en directo y en vídeo) así como la fiabilidad tras el consenso de los evaluadores previamente a la evaluación. Estas investigaciones han servido de fundamento para el desarrollo de A*lfamov,* la novedosa herramienta informática diseñada para la evaluación de las habilidades motrices básicas de niños y niñas de entre 6 y 12 años que se presenta como resultado final de esta tesis. Las publicaciones mencionadas y A*lfamov* podrán servir de base para una evaluación más sencilla, fiable y sistematizada de las habilidades motrices básicas tanto en la práctica como en la investigación.

Palabras clave: Hábitos saludables, Competencia motriz, Habilidades motrices básicas, Niños, Evaluación, Fiabilidad, App.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.2 RESUMO

O incremento de comportamentos sedentarios da nosa sociedade, así como as taxas elevadas alcanzadas de obesidade e a súa tendencia ao aumento nas últimas décadas, provoca que se continúen buscando e entendendo as súas causas. Neste sentido, reportáronse probas sobre o papel da competencia de habilidades motrices básicas na práctica de actividade física e, por conseguinte, na adquisición de estilos de vida saudables ao longo da vida. Por iso, é preocupante o descenso que sufriron os niveis de competencia de mencionadas habilidades nas últimas décadas. As medidas de intervención poderían enfocarse en tratar de incrementar a competencia motriz dos máis novos para conseguir un estilo de vida activo desde a infancia. Para poder implementar esta intervención é necesario coñecer o nivel de desenvolvemento de competencia motriz dos nenos e nenas a través da súa avaliación, para o que se describiron diferentes barreiras. Entre elas, o tempo necesario para levala a cabo xunto á escaseza do mesmo para o profesorado e o persoal sanitario de atención primaria, a falta de formación ou coñecementos entre os principais axentes implicados ou a necesidade de ferramentas adecuadas para levar a cabo a avaliación.

Nesta tese doutoral preséntanse, entre outros resultados adicionais, dous artigos científicos publicados e un en proceso de revisión que tratan de dar resposta a parte dos problemas mencionados. Concretamente, á fiabilidade da avaliación da competencia motriz, requisito imprescindible para realizar unha avaliación rigorosa.

A primeira publicación é unha revisión sistemática sobre a fiabilidade dun dos instrumentos máis utilizados para a avaliación da competencia motriz grosa, o Test of Gross Motor Development, mostrando que é unha ferramenta psicométrica robusta de avaliación. Os dous estudos restantes (un deles publicado), mostran os resultados de fiabilidade entre observadores de distintas características (novatos e expertos) con dous métodos de avaliación (en directo e en vídeo) así como a fiabilidade tras o consenso dos avaliadores previamente á avaliación. Estas investigacións serviron de fundamento para o

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

desenvolvemento de *Alfamov*, a nova ferramenta informática deseñada para a avaliación das habilidades motrices básicas de nenos e nenas de entre 6 e 12 anos que se presenta como resultado final desta tese. As publicacións mencionadas e *Alfamov* poderán servir de base para unha avaliación máis sinxela, fiable e sistematizada das habilidades motrices básicas tanto na práctica como na investigación.

Palabras chave: Hábitos saudables, Competencia motriz, Habilidades motrices básicas, Nenos, Avaliación, Fiabilidade, App.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.3 ABSTRACT

The increase in sedentary behaviours in our society, as well as the high rates of obesity and its increasing trend in the last decades, has led to a continuous search for and understanding of its causes. In this sense, evidence has been reported on the role of fundamental movement skills competence in the practice of physical activity and, consequently, in the acquisition of healthy lifelong lifestyles. It is alarming the decline in the levels of competence in these skills over the last decades. Intervention measures could focus on increasing Young people's motor competence to achieve an active lifestyle from childhood onwards. To implement this intervention it is necessary to know the level of development of children's motor competence, for which different barriers have been described: the time needed to carry it out together with the shortage of time for teachers and health care staff; the lack of training or knowledge about the assessment of motor competence among the main agents involved; or the need for adequate tools to carry out the assessment are some of them.

This doctoral thesis presents, among other additional results, two published scientific articles and one in the process of revision that attempt to provide an answer to some of the problems mentioned, specifically to the reliability of motor competence assessment, an essential requirement for a rigorous assessment.

The first publication is a systematic review of the reliability of one of the most widely used instruments for the assessment of gross motor competence, the Test of Gross Motor Development, showing that it is a robust psychometric assessment tool. The remaining two studies (one of them published) show the results of inter-rater reliability between raters with different backgrounds and experience (novice and expert) with two assessment methods (live and video) as well as inter-rater reliability after a pre-assessment consensus of the raters. This research has served as the basis for the development of *Alfamov*, the innovative IT tool designed for fundamental movement skills assessment in children aged 6-12 years, which is presented as the final result of this thesis. The aforementioned publications and *Alfamov* can serve as a basis for a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

more reliable, simple and systematised assessment of fundamental movement skills in practice and research.

Keywords: Healthy lifestyle, Motor competence, Fundamental movement skills, Children, Assessment, Reliability, App.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2 INTRODUCCIÓN

Los niños, niñas y adolescentes en ocasiones tienen que afrontar dificultades y obstáculos para poder crecer y desarrollarse sanos y sanas. A veces no pueden seguir una vida activa por falta de tiempo o lugar para jugar y practicar deporte, no siempre tienen acceso a una alimentación adecuada y, en ocasiones, tampoco cuentan con un entorno que favorezca su bienestar y descanso. Es necesario fomentar y promover estilos de vida saludables desde los primeros años de vida para conseguir una infancia, adolescencia y vida adulta sana.

2.1 ESTILOS DE VIDA SALUDABLES

2.1.1 Inactividad física, sedentarismo y obesidad: el problema

La actividad física (AF), es definida por la Organización Mundial de la Salud (OMS) como cualquier movimiento corporal producido por la musculatura esquelética que implique un gasto energético. A pesar de los importantes beneficios de la práctica de AF para la salud, los índices de práctica continúan siendo bajos[1]. A su vez, la inactividad física y el sedentarismo se han incrementado en los últimos años. Aunque en ocasiones se usen como sinónimos, por inactividad física se entiende el nivel insuficiente de práctica de AF que no cumple con las recomendaciones globales de la OMS necesarias para alcanzar una buena salud; por su lado, el sedentarismo se refiere a toda aquella actividad física por debajo del estado de reposo, definido con una intensidad inferior a 1.5 METs[2–4]. El aumento de estos hábitos poco saludables se debe principalmente a los cambios sociales, económicos y en el entorno físico de nuestra sociedad actual. La OMS ha alertado de que los niveles de inactividad física son alarmantes y se estima que suponen 54.000 millones de dólares de coste mundial en atención sanitaria directa cada año[2]. En su afán de combatir la inactividad física y los comportamientos

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sedentarios, la OMS ha elaborado el "*Global Action Plan on Physical Activity (GAPPA) 2018-2030"* donde se establece como objetivo principal la reducción global de los niveles de inactividad física en adolescentes y adultos en un 15% hasta 2030[2].

Los cambios en la sociedad, como el transporte, el mayor uso de nuevas tecnologías y la urbanización, que han provocado una transición a hábitos más sedentarios, han supuesto en algunos países niveles de inactividad física que llegan hasta el 70%[2]. Estos cambios afectan sobre todo a los países más desarrollados, razón que puede explicar por qué en los países de mayores ingresos los niveles de AF insuficiente son especialmente altos, entorno al 36.8%, en comparación, por ejemplo, a una prevalencia del 16.2% en países de bajos ingresos[5]. Más concretamente, en España, la prevalencia de AF insuficiente en adultos en 2016 se situaba en el sexo masculino entre el 20-29.9% y en el sexo femenino entre el 30-39.9%[5]. La prevalencia de práctica de AF insuficiente se incrementa si hablamos de niños y niñas entre 11 y 17 años, suponiendo en nuestro país en el año 2016 en torno al 69.8% en adolescentes varones y 83.8% en mujeres[6].

Cabe destacar que no seguir las recomendaciones de práctica de AF puede poner en riesgo la salud tanto inmediata como futura. De hecho, diversas evidencias científicas han señalado que los niveles elevados de sedentarismo e inactividad física se asocian con determinadas enfermedades no transmisibles; es decir, aquellas afecciones no causadas por una infección aguda, que tienen consecuencias para la salud a largo plazo y que generalmente requieren de atención y tratamiento[7]. En este sentido, se han relacionado los hábitos sedentarios con diferentes tipos de cáncer[8,9] así como con las enfermedades cardiovasculares[8,10] y con la diabetes[11]. También se ha asociado la inactividad física y sedentarismo con el mayor riesgo de obesidad[11] y mortalidad[12,13]. Al mismo tiempo, se ha evidenciado una relación inversa entre comportamientos sedentarios y la percepción de felicidad, el nivel subjetivo de salud, la calidad de vida[14] y la salud mental[15].

Por otro lado, la obesidad y el sobrepeso son considerados problemas relevantes de salud pública a nivel mundial, cuya prevalencia se ha triplicado desde los años 70. Cifras del año 2016

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

destacan que casi 2.000 millones de personas mayores de 18 años tenían sobrepeso en todo el mundo, de los cuales 650 millones eran obesos. En particular, en los más jóvenes, la prevalencia también se ha incrementado de forma alarmante. Se reportó que alrededor de 340 millones de niños, niñas y adolescentes de entre 5 a 19 años (en el año 2016) y 39 millones menores de 5 años (en el año 2020) sufrían obesidad o sobrepeso[16].

La obesidad suele iniciarse en la infancia y se asocia con distintos problemas de salud tanto en mencionada etapa como en la adolescencia y en la edad adulta. Por tanto, debería prevenirse desde un primer momento y no esperar a que exista para tratarla. Tal es el impacto de la obesidad en nuestra sociedad que la OMS la ha declarado "*epidemia del siglo XXI*" por los costes individuales, sociales y económicos que supone y supondrá en los próximos años[17].

Según el último Estudio de Alimentación, Actividad física, Desarrollo Infantil y Obesidad en España (ALADINO) publicado en el año 2019, 4 de cada 10 niños y niñas tienen exceso de peso en nuestro país. Respecto al informe previo (año 2015), la tendencia de sobrepeso y obesidad se mantiene en niveles altos de prevalencia[18], sin conseguir hacer frente a este problema a pesar de las numerosas iniciativas.

2.1.2 La actividad física como agente clave en la adquisición de un estilo de vida saludable

Existen datos consistentes que señalan los beneficios de la práctica regular de AF en la salud de la población[19]. Por eso se ha postulado la AF como un agente clave en la adquisición de un estilo de vida activo y en la prevención del sobrepeso y obesidad[20].

Se ha observado que la práctica regular de AF contribuye en la prevención de enfermedades no transmisibles como las enfermedades cardiovasculares, los accidentes cerebrovasculares, la diabetes y cáncer de mama o colon[21]. También actúa como factor protector en la prevención de la hipertensión, sobrepeso y obesidad[2], las enfermedades mentales[22] y en la reducción del riesgo de demencia[23]. De igual modo, mejora el bienestar y la calidad de vida[2,24]. Además, estos beneficios pueden observarse con "volúmenes" pequeños, es

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

decir, con una cantidad de práctica de AF relativamente baja[25], hecho que apoya su práctica en todas las edades y la necesidad de ser integrada en los diversos entornos cotidianos[2].

Sin embargo, y a pesar de los conocidos beneficios de la práctica de AF y de los esfuerzos por crear planes estratégicos de intervención, la OMS ha reportado que más de la mitad de la población sigue sin alcanzar unos índices mínimos de práctica de AF necesarios para que les aporte beneficios en la salud[2]. Esto dificulta alcanzar los objetivos de reducción de la inactividad física marcados por la OMS[2], ya que hay que luchar contra hábitos no saludables muy arraigados en las personas. De hecho, se estima que entre los años 2001 y 2016 se han mantenido prácticamente estables los niveles de prevalencia de insuficiente AF tanto en las personas adultas como en las más jóvenes[5,6]. Por otra parte, no solo son preocupantes las personas que nunca han practicado AF, sino aquellas que la practicaban y la abandonaron por falta de adherencia, hecho que a su vez tiende a relacionarse con la adquisición de estilos de vida sedentarios y poco saludables[26].

Entre las razones que han tratado de dar respuesta al abandono de la práctica regular de AF de niños, niñas y adolescentes se encuentran la falta de tiempo, el insuficiente disfrute y la presión generada por la competición, entrenadores y familias[27].

Otra de las razones de abandono más destacables es falta de habilidad o competencia motriz (CM), entendida como la capacidad de una persona para desenvolverse de forma eficiente en una amplia gama de actividades motrices de la vida diaria[28], o la baja percepción de competencia[27,29]. En relación con esto, Macarro Moreno et al.[26] en su estudio realizado con adolescentes mostraron que el 23.4% de los participantes manifestó haber abandonado la práctica de actividades físico-deportivas por no sentirse lo suficientemente hábiles. A su vez, el 24.4% expresó que el abandono se produjo "*por no ser tan bueno como pensaba*" y el 20.7% porque "*los demás eran mejores*". En la misma línea, Lopes et al.[30] observaron que los niños y niñas que presentaban altos niveles de CM disminuyeron en menor medida su práctica de AF 4 años después, en comparación con los que tenían un nivel de CM medio o bajo. Este hecho puede verse incrementado en el

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sexo femenino ya que, en términos generales, suelen tener una percepción más baja de CM, razón por la que tienen menor probabilidad de disfrutar de la práctica de AF[31] y mayor probabilidad de abandonarla[32,33].

Se han hecho esfuerzos para identificar las correlaciones de la AF con el objetivo de hacer frente a las tasas elevadas de inactividad física (asociadas a comportamientos más sedentarios y menos saludables) mediante intervenciones específicas. De este modo, las asociaciones identificadas de mayor interés están relacionadas con la CM real[34,35] y percibida[36,37] y con la autoestima y el autoconcepto[38]. Por esta razón, las recomendaciones para mantener la práctica de AF sitúan el énfasis en el incremento de la CM y disfrute desde edades tempranas[29]. Por tanto, se hipotetiza que si se aumenta la CM, aumentará también la percepción de la misma, el disfrute y la práctica de AF favoreciendo a su vez el descenso del sedentarismo.

2.2 ALFABETIZACIÓN MOTRIZ

Para tratar de incrementar los niveles y mantener la adherencia de práctica de AF, los objetivos no deben centrarse solamente en cambiar los hábitos poco saludables arraigados en la población, sino también en focalizar la atención en la infancia. De este modo, se conseguirá que las nuevas generaciones sean físicamente activas desde edades tempranas. Esta es la idea sobre la cual se sustenta, en términos generales, el concepto de Alfabetización Motriz.

2.2.1 Definición

La Alfabetización Motriz (AM) es un concepto holístico, universal e inclusivo que se define comúnmente como la motivación, la confianza, la competencia física, el conocimiento y la capacidad de comprensión para valorar y responsabilizarse de practicar AF a lo largo de la vida[39,40]. Se basa en la necesidad de desarrollar las habilidades, los conocimientos y comportamientos para convertirse en una persona físicamente activa y continuar siéndolo toda la vida.

El término se fundamenta en un símil que parte del concepto de alfabetización lingüística comparándola con la “destreza de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

movimiento"[41]. Del mismo modo que para el pleno desarrollo humano se considera esencial adquirir una adecuada alfabetización lingüística, es decir, las destrezas básicas de leer y escribir que constituyen la base del aprendizaje a lo largo de toda la vida[42], también es fundamental una adecuada AM que facilitará que una persona sea físicamente activa durante su vida.

Aunque Margaret Whitehead se considera la creadora del concepto de AM cuya definición ha ido evolucionando a lo largo de los años desde su propuesta en el año 1993[39,43], sigue existiendo un amplio debate sobre la propia la definición. Whitehead apuntaba que hay razones comprensibles para la existencia de diferentes interpretaciones de la AM. De hecho, tomando como referencia la propia definición, los aspectos personales, contextos sociales y ambientales que la conforman provoca que la AM requiera un enfoque específico para cada contexto[44]. Sin embargo, estos diferentes enfoques han dado lugar a múltiples definiciones que a menudo se interpretan de manera diferente[45]. Esto conduce a la preocupación de diferentes autores de que el concepto se confunda y se esté aplicando de forma inconsistente[41,46,47], limitando, por ejemplo, la definición de AM únicamente al área física, lo que se considera inadecuado como se explicará más adelante.

2.2.2 Evolución de los modelos conceptuales de desarrollo

En el 2008, Stodden et al.[48] propusieron un modelo (Figura 1) que hipotetizaba las relaciones entre la AF, la CM, la CM percibida y la condición física relacionada con la salud. Lo plantearon con la premisa de que "*las investigaciones anteriores no tuvieron en cuenta el papel dinámico y sinérgico que la competencia motriz desempeña en el inicio, mantenimiento y declive de la actividad física*". Este enfoque del desarrollo fue propuesto para abordar el papel potencial de la CM en la promoción de trayectorias positivas o negativas de: la percepción de CM, la AF y condición física relacionada con la salud. Este modelo propone sinergias recíprocas, que se refuerzan con el tiempo, de forma que si se desarrollan y se relacionan adecuadamente tendrán un efecto positivo en la salud. Si por el contrario, los elementos se relacionan de forma negativa, no se conseguirán estos efectos beneficiosos.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

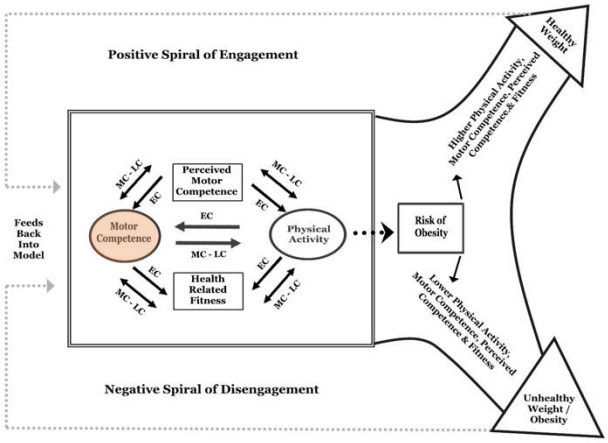


Figura 1 [1]**.** Modelo de desarrollo propuesto por Stodden et al.[48] *EC* Early childhood, *MD* middle childhood, *LC* late childhood.

Tras siete años de la publicación del modelo de desarrollo propuesto por Stodden et al. [48], Robinson et al.,[49] sugirieron otro modelo conceptual (Figura 2) acorde a los cambios y nuevas aportaciones de las investigaciones más recientes. Este modelo, basado en la evidencia científica publicada hasta el 2015, remarcó la existencia de una fuerte relación entre la CM y la AF en la infancia. También sugirió una fuerte asociación entre la CM y el estado de peso (inversa) y la condición física relacionada con la salud (directa) durante la infancia y adolescencia. Por otro lado, aunque la relación entre la CM y la percepción de la misma se mostró más variable con el paso del tiempo, se identificó la percepción de competencia como un mediador entre la CM y la AF, es decir, como agente influyente en la relación directa entre CM y AF. Robinson et al., concluyen su trabajo indicando que el desarrollo físico y psicológico de los niños y niñas es un laberinto complejo en el que confluyen factores

[1] Extraído con permiso de Springer Nature: Springer, Sports Medicine, bajo la licencia Creative Commons Attribution 4.0 (que permite reproducir esta publicación siempre que se atribuya a la obra). *Through the looking glass: a systematic review of longitudinal evidence, providing new insight for motor competence and health.* Barnett et al. Copyright (2022).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

biológicos, ambientales, psicosociales y conductuales que evolucionan de forma sinérgica a lo largo del tiempo.

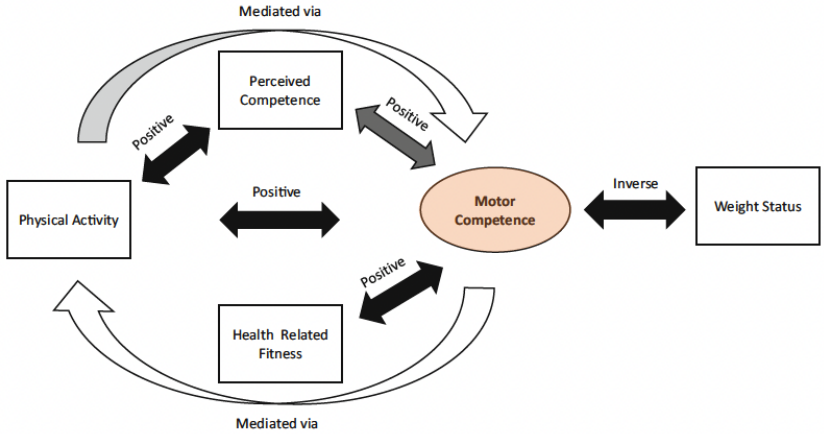


Figura 2 [2]. Modelo conceptual propuesto por Robinson et al.[49] basado en la evidencia científica sobre la relación de la CM y variables relacionadas con la salud. *Las flechas negras indican que se ha comprobado con evidencia: relación consistente; las flechas grises oscuro que se ha comprobado de forma moderada, relación variable; las flechas grises claro que se ha comprobado parcialmente, que existen algunas pruebas; y la flecha blanca indica que se ha comprobado de manera limitada.*

En el año 2018, Hulteen et al.[50] propusieron otro modelo (Figura 3) que ilustra el desarrollo de las habilidades motrices desde los movimientos reflejos hasta las habilidades especializadas. Para ello se basaron en los modelos de Stodden et al.,[48] y de Robinson et al.,[49] y en la evidencia previa al 2018. Así, sugirieron una clasificación de habilidades motrices (a las que denominaron "*foundational movement skills*") esenciales para la participación en AF. Al mismo tiempo, reconocen la posible influencia de la diversidad cultural y geográfica y el papel sinérgico y recíproco de factores físicos y psicológicos, como la percepción de competencia, el peso o la autoestima, en el desarrollo de las habilidades motrices.

[2] Extraído con permiso de Springer Nature: Springer, Sports Medicine. *Motor competence and its effect on positive developmental trajectories of health*, Leah E Robinson et al. Copyright (2015). Licencia número: 5390160244369.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

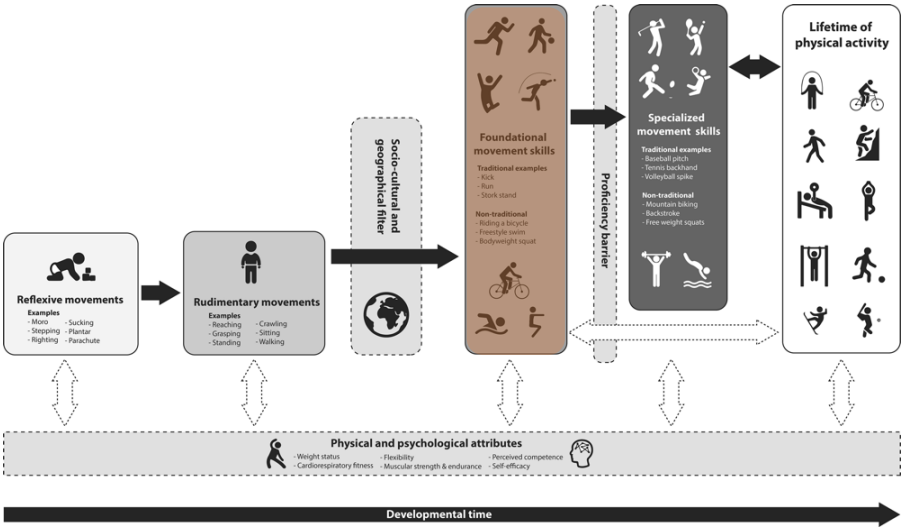


Figura 3 [3]**.** Modelo de desarrollo de habilidades motrices básicas para la práctica de actividad física a lo largo de la vida propuesto por Hulteen et al[50]. **Las flechas negras indican patrones hipotetizados previamente y las blancas y los recuadros de puntos las nuevas vías/componentes hipotéticos propuestos por este modelo.*

No es hasta 2019 cuando Cairney et al.[51] publicaron el primer modelo conceptual donde se incluye la AM (Figura 4). Su objetivo fue posicionar a la AM como un determinante de salud y enfermedad apoyándose para ello en la evidencia científica que lo respaldaba. Este modelo sugirió que la AM debe considerarse como un constructo dinámico en el tiempo donde confluyen además de la dimensión física y psicológica, la social y cognitiva. Como se puede observar en la Figura 4, para conectar la AM y la AF se han utilizado dos flechas en sentido opuesto con el afán de enfatizar la importancia de la relación recíproca entre ambas. Es decir, Cairney et al.[51] consideran que la AM es esencial para la participación en AF, pero a su vez, para alcanzar la AM es necesaria la práctica de AF (tanto organizada como libre). Siguiendo la propuesta de Hulteen et al.[50], este modelo también

[3] Extraído con permiso de Springer Nature: Springer, Sports Medicine. *Development of foundational movement skills: a conceptual model for physical activity across the lifespan*, Ryan M Hulteen et al. Copyright (2018). Licencia número: 5390160456975.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

incluye los factores individuales (p. ej., raza, sexo o personalidad) y ambientales (p. ej., climatología o el lugar donde vivas) como atributos moderadores de la relación entre la AM y la AF. Si estos elementos siguen una trayectoria positiva a lo largo de la vida tendrán efectos positivos en la salud.

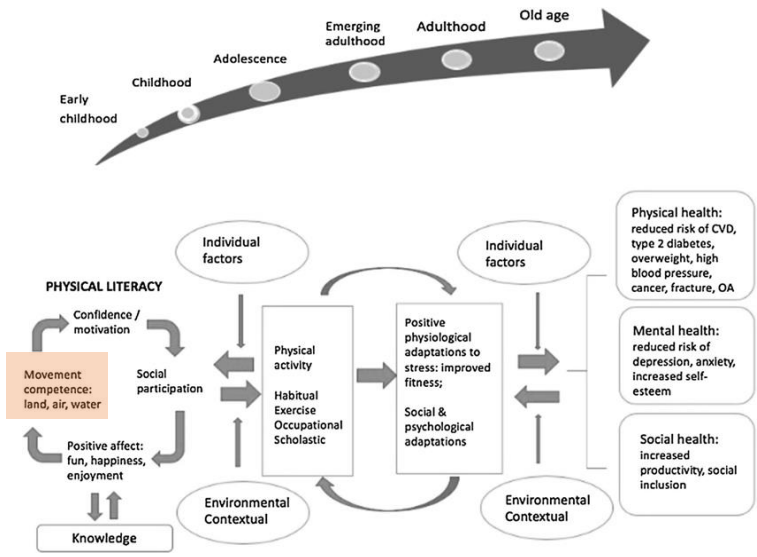


Figura 4 [4]. Modelo conceptual propuesto por Cairney et al.[51] que vincula la AM, la AF y la salud. *CVD* cardiovascular disease (enfermerdades cardiovasculares), *OA* osteoarthritis (osteoartritis).

En una revisión sistemática reciente, Barnett et al.[52] revisaron los dos modelos de Stodden et al.[48] y Robinson et al.[49] presentados anteriormente y realizaron una propuesta (Figura 5) recopilando la creciente evidencia publicada en torno a la CM desde el 2015. Encontraron pruebas sólidas (señalizado en la Figura 5 con flechas negras) de la relación negativa entre la CM y el peso y sobre la relación positiva entre la CM y la condición física relacionada con la salud. En cuanto a la relación entre la CM y la AF, observaron

[4] Extraído con permiso de Springer Nature: Springer, Sports Medicine. *Physical Literacy, physical activity and health: toward an evidence – Informed conceptual model*, John Cairney et al. Copyright (2019). Licencia número: 5390160596491.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

evidencia indeterminada. Por otro lado, no encontraron suficientes pruebas de la asociación entre la CM y la CM percibida. Debido a ello, estos autores han destacado la necesidad de seguir buscando evidencias científicas que aporten mayor rigurosidad a las encontradas hasta el momento.

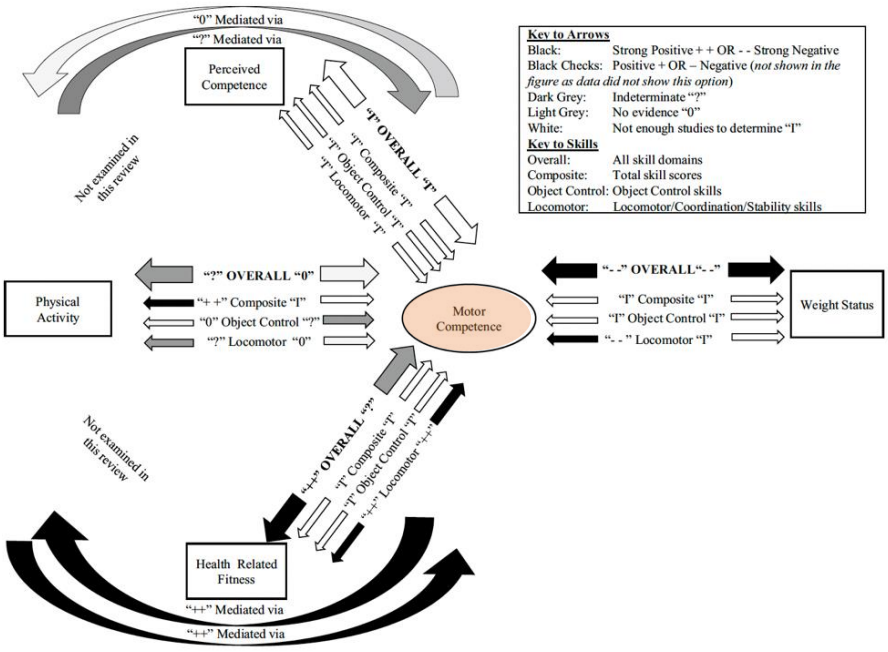


Figura 5 [5]**.** Modelo propuesto por Barnett et al.[52].

Por otro lado, centrándonos más concretamente en el desarrollo de la AM, se puede tomar como referencia para comprender el concepto el modelo australiano de AM[53]. Este modelo (Figura 6) se compone de cuatro piezas no jerarquizadas incluyendo las

[5] Extraído con permiso de Springer Nature: Springer, Sports Medicine, bajo la licencia Creative Commons Attribution 4.0 (que permite reproducir esta publicación siempre que se atribuya a la obra). *Through the looking glass: a systematic review of longitudinal evidence, providing new insight for motor competence and health.* Barnett et al. Copyright (2022).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

capacidades físicas, psicológicas, sociales y cognitivas. Estos elementos, denominados *dominios* tienen el mismo peso relativo en la AM y están relacionados de modo que el desarrollo de cada uno favorecería a los demás. Cada dominio está formado por elementos clave que contribuyen de forma conjunta en el completo desarrollo de la AM.



Figura 6[6]. Modelo australiano de AM. (Sports Australia 2019).

El interés y énfasis en la CM (*véase parte naranja de las figuras 1-6)* como elemento clave en la relación con distintos indicadores de salud es creciente[54]. No obstante, la revisión sistemática de Barnett et al.[52] concluye que siguen siendo necesarios más estudios (sobre todo longitudinales en la infancia y adolescencia) para probar con mayor solidez las relaciones expuestas en los modelos.

2.2.3 Dominios de la alfabetización motriz

Se ha propuesto que el desarrollo de la AM implica un aprendizaje holístico a través de cada uno de los elementos que la componen[53]. Como se ha mencionado anteriormente, estos elementos suelen recibir el nombre de "dominios" o "componentes", y aunque

[6] Extraído con permiso de Sports Australia bajo la licencia Creative Commons (que permite reproducir esta publicación siempre que se atribuya a la obra). *Australian Physical Literacy Framework.* Keegan et al. Copyright (2019).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

los distintos autores difieren ligeramente en la denominación de cada uno de ellos, la mayoría concuerdan que existen al menos cuatro dominios básicos: Físico, Psicológico o Afectivo, Cognitivo y Social[45].

2.2.3.1 Dominio Físico

Engloba la CM, el control corporal y la condición física como elementos necesarios para llevar a cabo los movimientos que requieren las diferentes situaciones o entornos de la vida diaria[53]. Varias definiciones de AM incluyen las habilidades motrices básicas (HMB) como el componente principal del dominio físico[55–58]. Además, las que no incluyen las HMB en la propia definición las utilizan para explicar la base de este dominio.

Sin embargo, es importante recordar que no es adecuado limitar la definición de AM únicamente al desarrollo de las HMB[44,59], porque esto podría provocar que se desvinculase la AM de su significado central y se dejase ausentes a otros dominios importantes[60].

2.2.3.2 Dominio Psicológico

Este dominio también se ha denominado "Afectivo" por algunos autores[55,61–65] y se fundamenta en los sentimientos, actitudes y emociones hacia la AF[53]. Se basa en el desarrollo de la autoestima, la confianza, la motivación y el disfrute de las diferentes experiencias del movimiento que repercutirán en el deseo de ser físicamente activo. Es decir, las personas tienen que ser capaces de comprender y gestionar las respuestas emocionales tales como la empatía y la sensibilidad, así como las respuestas físicas como el dolor y la fatiga al practicar AF. Esto implica persistir en la práctica de AF independientemente de las dificultades, retos y fracasos que pueda conllevar, con la convicción de que la mejora llegará con el aprendizaje y esfuerzo[53].

2.2.3.3 Dominio Cognitivo

Este dominio está centrado en la adquisición individual de los conocimientos y la comprensión necesaria para la práctica de AF[53]. Una persona tiene que saber cómo, cuándo y por qué moverse y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ser consciente de los beneficios para la salud de un estilo de vida activo que incluya la AF. Por otro lado, este dominio también comprende la conciencia de cada individuo de sus propias capacidades y qué hacer para mejorar tanto las capacidades físicas como su estado de salud y bienestar[44].

2.2.3.4 Dominio Social

Es un dominio que se focaliza en el desarrollo de las habilidades sociales que permitan la interacción con otras personas en torno al movimiento y la AF[53]. Ejemplos de habilidades sociales son la colaboración, la comunicación, el juego limpio, el liderazgo, entre otras que ayudarían a disfrutar de la participación con los demás en distintas actividades. Algunos autores no han incluido este dominio en su clasificación al considerar que los elementos que lo conforman (al menos la comunicación fluida y la interacción empática con los demás) se deben incorporar dentro del dominio psicológico o afectivo [44].

Otros autores han incluido el *dominio de comportamiento* (*behavioural domain*) para referirse a los comportamientos e intereses que conducirán al compromiso de participación en hábitos de AF de por vida, responsabilidad que debe ser asumida por cada persona individualmente[55,64,65].

2.3 COMPETENCIA MOTRIZ

Tras el análisis de la evolución de los modelos conceptuales de desarrollo, del propio concepto de AM y de los dominios que la conforman parece evidente el énfasis puesto en la CM como uno de los elementos claves para alcanzar la AM a lo largo de la vida. Es por esto que Young et al.[66] consideraron la mejora de la CM como elemento necesario para incrementar la participación en juegos, deportes y en la actividad física general, así como uno de los objetivos imprescindibles para alcanzar la AM.

La CM se ha definido como *"la capacidad de una persona para realizar diferentes actividades que impliquen motricidad, incluidas la coordinación de las habilidades motrices finas y gruesas, necesarias para desenvolverse en las tareas de la vida cotidiana"* [28].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Existen evidencias que sostienen la relación entre la CM y los hábitos de vida saludables[67,68], por lo que en los últimos años la CM ha sido objeto de un interés cada vez mayor como mecanismo y estrategia potencial para controlar o atenuar problemas de salud como el sobrepeso y la obesidad[69,70]. A su vez, la CM gruesa es caracterizada por utilizar los músculos grandes para realizar los movimientos requeridos para actividades de la vida cotidiana, así como para actividades deportivas específicas[71]. Puede tener un importante rol en el crecimiento físico, el desarrollo psicomotor y también generar oportunidades para seguir un estilo de vida activo a lo largo de toda la vida[35]. La CM gruesa se suele describir como la competencia en las HMB, y su desarrollo ocurre durante la infancia[71].

2.4 HABILIDADES MOTRICES BÁSICAS

2.4.1 Definición

Las HMB (*fundamental movement skills* en inglés) se definen como una serie de movimientos básicos y organizados que implican la combinación de patrones de movimientos de dos o más partes del cuerpo[72], y son consideradas como los pilares del movimiento. Corlett y Mandigo[73] las sitúan como el primer nivel del dominio Físico necesario para alcanzar la AM. Retomando el símil que comparaba la alfabetización del lenguaje con la AM, las HMB pueden considerarse análogas a las letras del abecedario "*el ABC del movimiento*"[74]. Al igual que las letras son la base para aprender palabras, que luego permitirán construir frases para poder comunicarse, las HMB funcionarían como la raíz para la adquisición de otras habilidades específicas y más complejas[71]. Se asume que si no se aprenden las competencias básicas de comprensión de las letras se producirán problemas en el desarrollo normal del lenguaje. Del mismo modo, si no se adquieren y desarrollan las HMB en la infancia puede comprometerse el desarrollo motor completo que afectará en la realización de actividades que impliquen la motricidad no solo en el presente sino a lo largo de toda la vida.

Las HMB se clasifican en habilidades Locomotrices, Manipulativas y de Estabilidad (*véase apartado siguiente*). Si bien su

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

grado de desarrollo varía en función de la habilidad concreta y por la influencia de otros factores (como el sexo o la edad), generalmente abarca desde los 3 a los 11 aproximadamente[71].

La revisión sistemática realizada por Barnett et al.[75] sobre los determinantes de la competencia en las HMB concluyó que la edad (creciente), el peso (saludable), el sexo (masculino) y el nivel socio-económico (más alto) son indicadores de la CM. Dichos indicadores deben tenerse en cuenta en las intervenciones para mejorar la CM, que deberán orientarse hacia los grupos más desfavorecidos. Particularmente, las diferencias en el sexo son evidentes sobre todo para las habilidades manipulativas[70,76]. Esto se deben a factores biológicos y evolutivos que diferencian a los niños y las niñas y provocan que en habilidades complejas como los lanzamientos que requieren una velocidad y fuerza considerables los niños obtengan mejores resultados (en cuanto a distancia o rapidez o de lanzamiento)[77,78]. Por otro lado, las diferencias en el sexo también son debidas a factores socioculturales, relacionados con los estereotipos de género que provocan que las niñas tengan menos estímulos y oportunidades para practicar las habilidades manipulativas[75].

2.4.2 Clasificación

Los niños y niñas son capaces de explorar el potencial de sus cuerpos al desplazarse (gracias a las habilidades locomotrices), tienen control sobre su musculatura en oposición a la gravedad (gracias a las habilidades de estabilidad) y adquieren cada vez una mayor capacidad para establecer un contacto de forma controlada y precisa con los objetos del entorno (gracias a las habilidades manipulativas)[71]. Según eso las HMB se han clasificado en tres grupos: locomotrices, manipulativas o de control de objetos y de estabilidad.

Las habilidades locomotrices, (por ejemplo correr o trepar) son aquellas que permiten la traslación del cuerpo en el espacio y, por tanto, desplazarse de un punto a otro.

Las habilidades manipulativas, también denominadas habilidades de control de objetos, (como lanzar o recepcionar) son aquellas que implican, el control y/o la manipulación de los objetos.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Y las habilidades de estabilidad (como rodar o mantener el equilibrio) se caracterizan por el mantenimiento de la estabilidad postural durante la estática o la dinámica del movimiento[71,79].

En la Tabla 1 se pueden observar ejemplos de las HMB más comunes de cada grupo.

Tradicionalmente, las habilidades de estabilidad se clasificaban como subyacentes a las locomotrices[54], pero las evidencias científicas más recientes apoyan la inclusión de este grupo de HMB de forma independiente[80,81].

Tabla 1. Habilidades motrices básicas más comunes. [Elaboración propia]

Locomotrices	*Manipulativas*	*De Estabilidad*
Carrera	Lanzamiento	Equilibrio estático
Salto con un pie	Recepción	Equilibrio dinámico
Salto vertical	Patada	Voltereta hacia delante
Galope	Bote estático	Voltereta hacia atrás
Skip	Golpeo con raqueta	Rodar
Trepar	Bateo	
Salto horizontal		
Cambio de dirección		
Desplazamiento lateral		

2.5 EVALUACIÓN DE LAS HABILIDADES MOTRICES BÁSICAS

Las herramientas de evaluación de las HMB pueden clasificarse en tres grupos: a) las pruebas orientadas al producto, que evalúan el resultado final de la ejecución de una o más habilidades (p. ej., la distancia recorrida realizando un salto horizontal o de proyección en un lanzamiento); b) las pruebas orientadas al proceso**,** que evalúan la calidad del movimiento teniendo en cuenta la presencia o ausencia de ciertos patrones de movimiento (p. ej., la acción del brazo dominante en un lanzamiento o el movimiento de los brazos en oposición a las piernas en la carrera)[82]; y c) las pruebas híbridas o mixtas que evalúan al mismo tiempo el resultado y la presencia o ausencia de patrones de movimiento[83,84].

En la práctica, la elección y utilización de una batería u otra dependerá del propósito de la evaluación, así como del tiempo disponible, la formación de los evaluadores, el esfuerzo que pueda realizarse y coste que se esté dispuesto o se pueda asumir[82]. Sin

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

embargo, algunos expertos han mostrado su predilección por las baterías orientadas al proceso frente a las orientadas al producto, ya que las primeras aportan mayor información sobre el desarrollo[81] y sobre cómo mejorar los aspectos específicos de una habilidad[76,83].

2.5.1 Importancia del conocimiento y seguimiento del grado de desarrollo de las HMB

El desarrollo de las HMB es esencial para conseguir realizar las tareas básicas de la vida cotidiana y también para alcanzar la competencia en diferentes juegos o deportes propios de nuestra sociedad[50,79]. Se ha establecido que la mayoría de los niños y niñas de 10 años deberían mostrar competencia o maestría (entendiéndose como la ejecución correcta) en la mayoría de las HMB[71]. Sin embargo, en países como Reino Unido e Irlanda, se ha observado que la competencia motriz de los niños y niñas es preocupantemente baja y no es suficiente para alcanzar una vida físicamente activa[68].

Se creía que las HMB se aprenden y desarrollan de forma natural[85], sin embargo, tras estudiar los niveles de competencia de las HMB y realizar un seguimiento a lo largo de los años se observó un descenso en las últimas décadas[86] que, de ser cierta la premisa del desarrollo innato de las HMB, esto no ocurriría. Por tanto, los bajos niveles de competencia de HMB constituyen una preocupación a nivel mundial[76]. Aunque existen distintos factores individuales (como la edad o el peso, la motivación, o la coordinación), ambientales (como la climatología o el entorno donde se lleve a cabo una actividad) y propios de cada habilidad (como la velocidad o precisión necesaria para conseguir éxito) que pueden afectar en la competencia de cada una, se entiende que algunos de los factores pueden ser modificados por el profesorado, el personal sanitario y/o las propias familias para evitar que afecten negativamente en la competencia de las HMB[71]. Para ello, es necesario evaluar las HMB para detectar qué posibles problemas o condicionantes interfieren en el desarrollo de las mismas y según esto poder promoverlas y mejorarlas.

Por otra parte, es preocupante el declive secular de la motricidad gruesa[86,87] y la repercusión que esto puede tener en la salud de la futura población adulta. A este respecto, se ha observado una relación

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

directa entre el desarrollo de las HMB durante la infancia, la práctica de AF en años posteriores[30,49,88–92] y la mejor condición física[93,94]. También se observó una relación inversa entre el desarrollo de las HMB y el Índice de Masa Corporal (IMC), de modo que a mayor IMC (lo que puede indicar mayor porcentaje de masa grasa) peores resultados de competencia en las HMB[90,91,95]. Además, se ha comunicado que los niños, niñas y adolescentes con mayor competencia en las HMB tenían también mejor percepción de la CM propia[36,96,97]. Este hecho puede ser importante porque aquellos jóvenes que se sienten competentes al participar en juegos y/o deportes tienen más probabilidad de querer hacerlo y disfrutar de esta participación, lo que hará que se sientan motivados para seguir practicando AF [98]. Por otro lado, se ha mostrado que el dominio de las HMB en los más jóvenes contribuye a mejores resultados académicos[99,100] y al desarrollo de otras habilidades cognitivas como la memoria, la atención y el tiempo de reacción[100–102].

Por tanto, diversos estudios reivindican la importancia de la competencia de las HMB como un buen predictor de salud, con influencia en el desarrollo físico, psicológico, cognitivo y social[101]. En relación con el desarrollo físico, se ha mostrado además que los niños y niñas con una competencia en las HMB por debajo de los estándares para su edad muestran dificultades para mejorar el rendimiento en habilidades más complejas, incluso tras un período de intervención educativa específica de dichas habilidades[103]. De esta forma, se justifica la evaluación sistemática de las HMB para conocer si el grado de desarrollo se corresponde con la edad biológica del niño o niña y para intervenir de forma temprana en caso de que esto no ocurra, ya que el dominio de las HMB parece ser una condición necesaria para el completo desarrollo motor.

2.5.2 La evaluación sistemática de las habilidades motrices básicas: una carencia a solventar

A pesar de la amplia variedad de instrumentos existentes y propuestos para la evaluación de las HMB, siendo 57 las pruebas identificadas recientemente[84], seguimos careciendo de un consenso sobre cuál es más adecuado en un contexto determinado[104]. En la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 2 se recogen algunas de las pruebas más utilizadas de las que se han reportado valores de fiabilidad[84,105].

En una revisión sistemática reciente[105], sobre la fiabilidad y validez de las herramientas de evaluación de las HMB, se mostró al Test of Gross Motor Development (TGMD) como uno de los instrumentos más válidos y fiables. Sin embargo, hay que tener en cuenta que la idoneidad de cada herramienta puede variar dependiendo del contexto en el que vaya a ser utilizada[106]. Considerando que la infancia representa el período más importante en el desarrollo de las HMB[72], se ha propuesto el colegio como el entorno clave para favorecer y evaluar su desarrollo. No obstante, los docentes han manifestado tener falta de conocimientos y formación[107]. En consecuencia, la evaluación de las HMB en los centros escolares puede ser compleja. Además, la saturación de instrumentos existentes para realizar dicha evaluación puede dificultar al profesorado en la elección del más adecuado. Por tanto, la evaluación de las HMB en los centros educativos se está descuidando a pesar de que el profesorado reconoce la importancia de hacerlo[108].

Aunque es probable que los instrumentos de evaluación actuales no sean completamente viables para su uso sistemático en los colegios[105] por el tiempo necesario, material requerido y los costes asociados[83], un estudio realizado con más de 800 profesores de educación primaria mostró que la falta de formación entorno a la competencia motriz era la causa principal por la que los docentes no evaluaban las HMB[109] .

En lo que respecta a la evaluación de la CM por parte del personal sanitario, realizada en los controles periódicos dentro del Programa de Salud Infantil de nuestro país, se utilizan instrumentos como el Test de Denver[110] o Haizae-Llevant[111]. Se ha propuesto que se realice la valoración de la CM en los controles de salud hasta aproximadamente los 5 años[112], quedando sin valorar años críticos para la consecución de una adecuada CM teniendo en cuenta que el desarrollo de las HMB se extiende hasta los 11 años aproximadamente [71].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 2. Principales test de evaluación de las habilidades motrices básicas. [Elaboración propia]

Nombre del test	*Habilidades evaluadas*	*Primer autor (año)*
Athletic Skills Track (AST)[a]	HL, HM	Hoeboer 2016[113]
Bruininks-Oseretsky Test of Motor Proficiency (BOTMP)[a]	HL, HM, HE	Bruininks, 2005[114]
Körperkoordinationstest für Kinder (KTK)[a]	HL, HE	Kiphard, 2007[115]
Motor Competence Assessment (MCA)[a]	HL, HM, HE	Luz, 2016[116]
Movement Assessment Battery for children (Movement-ABC)[a]	HL, HM, HE	Henderson y Sugden, 1992[28]
Get skilled Get Active[b]	HL, HM, HE	NSW, department of education and Training, 2000[117]
Objectives-Based Motor-Skill Assessment Instrument[b]	HL, HM	Ulrich, 1983[118]
Test of Gross Motor Development (TGMD)[b]	HL, HM	Ulrich, 2000[119]
Victorian HMB Assessmen Instrument[b]	HL, HM	Victoria, 1996[120]
Canadian Agility and Movement Skill Assessment (CAMSA)[c]	HL, HM	Longmuir, 2015[121]

[a] Test orientados al producto, [b] test orientados al proceso, [c] test híbridos
HL habilidades locomotrices, *HM* habilidades manipulativas, *HE* habilidades de estabilidad

Concretamente, en la Comunidad Autónoma de Galicia, la guía de salud infantil establece que la valoración del desarrollo psicomotor se hará de forma directa en lactantes y a través de una entrevista a sus familias en niños de mayor edad[122]. Esto puede dar lugar a una pérdida de información importante sobre el desarrollo real porque, aunque se ha mostrado que la percepción de los padres y madres puede servir como indicador de la CM real de los niños y niñas[123], también se ha reportado que las familias pueden sobreestimar el desarrollo psicomotor de sus hijos[124]. Por tanto, parece necesaria la evaluación de la CM en niños mayores de un año para asegurarse de conocer su competencia real.

Por otro lado, los planes de salud y controles del niño sano no son homogéneos entre comunidades autónomas ni entre países, lo que podría implicar un problema añadido en la estandarización de la evaluación de las HMB. En línea con lo anterior, la Academia Americana de Pediatría destaca en sus directrices para la supervisión sanitaria de niños y adolescentes la necesidad de conocer el desarrollo

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de habilidades motrices gruesas tales como correr, trepar, saltar o golpear un balón desde la primera infancia[125]. No obstante, la falta de homogeneidad hace que, por ejemplo, en nuestro país no se considere en los controles de salud la valoración de estas habilidades.

Además, los profesionales de la salud recomiendan la participación de las escuelas para trabajar de forma cooperada en la mejora de la salud y bienestar infantil y proponen a los colegios como escenario para la evaluación de las HMB[126]. Sin embargo, no existe una coordinación entre el sistema de salud y el educativo en este ámbito, necesaria para la promoción coordinada de hábitos saludables en general, y para la mejora de la competencia motriz en concreto.

En definitiva, la evaluación de las HMB sigue siendo un problema por solventar y, debido a ello, no se está realizando de forma adecuada por falta de tiempo, formación y conocimientos para llevarla a cabo o herramientas apropiadas.

2.5.3 Test of Gross Motor Development (TGMD)

El Test of Gross Motor Development (TGMD) (Test del Desarrollo Motor Grueso) y sus variantes (Test of Gross Motor Development – Second Edition [TGMD-2] y Test of Gross Motor Development – Third Edition [TGMD-3]) es un test orientado al proceso utilizado para determinar la CM gruesa a través de la evaluación de las HMB de niños y niñas de edades comprendidas entre 3 años y 10 años y 11 meses de edad[119,127,128].

Esta herramienta se divide en dos subescalas: a) la locomotriz y b) la de control de objetos. La subescala locomotriz incluye las HMB que requieren coordinación fluida de los movimientos del cuerpo para desplazarse (la carrera o el galope), y la subescala de control de objetos incluye las habilidades que muestran la competencia de los movimientos de lanzamiento, golpeo y recepción (como el bote estático de una pelota o la patada). En la Tabla 3 se presenta la clasificación de habilidades que componen el TGMD en sus tres versiones[119,127,128]. Las modificaciones de la segunda y tercera versión fueron realizadas de acuerdo a revisiones sistemáticas y recomendaciones de expertos según fueron obteniéndose nuevas evidencias científicas[128].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 3. Evolución de las habilidades locomotrices y de control de objetos en las versiones del Test of Gross Motor Development. [Elaboración propia]

TGMD[127]	*TGMD-2*[119]	*TGMD-3*[128]
Carrera	Carrera	Carrera
Galope	Galope	Galope
Salto con un pie	Salto con un pie	Salto con un pie
Salto horizontal	Salto horizontal	Salto horizontal
Desplazamiento lateral	Desplazamiento lateral	Desplazamiento lateral
Zancada	Zancada	--
Skip	--	Skip
Bateo	Bateo	Bateo
Bote estático	Bote estático	Bote estático
Recepción	Recepción	Recepción
Patada	Patada	Patada
Lanz. por encima	Lanz. por encima	Lanz. por encima
--	Lanz. por debajo rodando*	Lanz. por debajo*
--	--	Golpeo con raqueta

Lanz lanzamiento
* Habilidad que sufrió modificaciones de la segunda a la tercera versión del TGMD.

Cada habilidad incluye entre 3 y 5 criterios de rendimiento (relativos a los patrones motrices de cada HMB), que se puntúan con 1 o 0 en función de su presencia o ausencia, respectivamente. Estos criterios evalúan las características cualitativas del movimiento. Por ejemplo, en la carrera se tiene en cuenta si "*los brazos se mueven en oposición a las piernas*" o si "*el contacto del pie se reduce al talón o a la punta*"; y en el bote estático si "*contacta el balón con la mano a la altura de la cintura*" o si "*empuja el balón con la yema de los dedos sin dar golpes ni usar las palmas*".

Se puntuarán dos intentos de cada habilidad de los niños y niñas que la estén ejecutando tras la realización de un primer ensayo (que no se puntúa). La suma de los dos intentos realizados dará lugar a la puntuación de cada habilidad. A su vez, la suma de las habilidades de cada subescala dará lugar a la puntuación de la subescala locomotriz y de control de objetos, y la suma de ambas genera la puntuación global del TGMD. En el caso del TGMD-3, la última versión, puede obtenerse una puntuación máxima en las habilidades locomotrices de 46 puntos y de 54 puntos en las de control de objetos, siendo el total un máximo de 100 puntos (Tabla 4).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 4. Habilidades, criterios de rendimiento y puntuaciones del TGMD-3. [Elaboración propia]

Habilidades locomotrices			Habilidades de control de objetos		
Habilidades	**Núm. de criterios**	**Puntuación máxima**	**Habilidades**	**Núm. de criterios**	**Puntuación máxima**
Carrera	4	8	Bateo	5	10
Galope	4	8	Bote estático	3	6
Salto con un pie	4	8	Recepción	3	6
Skip	3	6	Patada	4	8
Salto horizontal	4	8	Lanz. encima	4	8
Desplazamiento lateral	4	8	Golpeo con raqueta	4	8
			Lanz. debajo	4	8
Puntuación subescalas		46			54

Núm. número, *Lanz.* lanzamiento

Por otra parte, en sus versiones se pueden encontrar los datos normativos actualizados de desarrollo de las habilidades motrices gruesas de niños y niñas de entre 3 y 10 años y 11 meses de edad para la población estadounidense.

Por tanto, entre los objetivos de uso del TGMD se encuentran[128]:

1) Evaluar el progreso individual de desarrollo las HMB.
2) Detectar los jóvenes que se encuentran significativamente por debajo de los datos normativos de desarrollo.
3) Planificar programas de intervención y evaluar su éxito.

Este test ha mostrado los resultados más favorables en términos de validez y fiabilidad en comparación a otras herramientas[105] y es uno de los más utilizados en la evaluación de las HMB por su bajo coste y viabilidad[106,119,127,128]. Por estas razones, y su utilidad con fines de investigación[128], ha sido elegido para la realización de los estudios incluidos en esta tesis doctoral.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.5.4 La fiabilidad como requisito imprescindible en la evaluación de las HMB

La fiabilidad es un requisito previo a la utilización de cualquier instrumento de medición o herramienta de evaluación en el ámbito clínico, educativo o de investigación[129,130]. Se define como el grado de consistencia de reproducción de resultados en el tiempo y el espacio y entre distintos observadores o evaluadores, y refleja la cantidad de error tanto aleatorio como sistemático implícito a cualquier medición[130,131]. Por otra parte estaría la validez, que se define como la capacidad de una herramienta de medición para reflejar lo que está diseñada para medir[132].

En lo que se refiere a la investigación, la fiabilidad y validez determinan la calidad y el rigor de la investigación, es decir, la adecuación metodológica, la idoneidad de la técnica o instrumento de recogida de datos, y si el análisis de los datos es el adecuado para medir las variables de estudio; por tanto, es un indicador de veracidad de los resultados de la investigación[132,133]. Se ha propuesto el Coeficiente de Correlación Intraclase (CCI) como la medida recomendable para el cálculo de la fiabilidad por reflejar tanto el grado de correlación como de concordancia entre las mediciones[129].

En relación con esto, para obtener datos rigurosos en la evaluación del desarrollo de las HMB es necesario que los instrumentos sean lo más fiables posible. Sin embargo, la evaluación de la CM utilizando pruebas orientadas al proceso (como el TGMD) puede ser compleja y requerir de un nivel elevado de conocimientos y formación específica por parte de los evaluadores para que logren la precisión adecuada[81,106].

Más concretamente, el TGMD, en su segunda y tercera versión, ha mostrado ser una herramienta robusta para la evaluación de las HMB presentando resultados de fiabilidad superiores a los de otros instrumentos. No obstante, se señala que la fiabilidad puede verse afectada por las distintas interpretaciones de cada evaluador a la hora de entender y puntuar los criterios de rendimiento de las HMB[105].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.6 JUSTIFICACIÓN

Existe un problema real de escasez de práctica de actividad física regular tanto en los más jóvenes como en la población adulta que influye negativamente no solo en el bienestar y calidad de vida individual, sino también en la comunidad, afectando al sistema sanitario y desarrollo económico[2]. A pesar de los esfuerzos por reducir los niveles de inactividad física y comportamientos sedentarios, estos se han incrementado, sobre todo en los países de renta alta, a lo largo del tiempo[5]. Paralelamente, la prevalencia de sobrepeso y obesidad se mantuvo alta en los últimos años tras haberse triplicado desde los años 70[16]. Por tanto, si no se ha conseguido reducir estos niveles a pesar de las numerosas investigaciones que proponen iniciativas, recomendaciones y programas de intervención, es necesario explorar otras vías de abordaje del problema.

En los últimos años se ha puesto el foco de atención en la importancia de la CM como mecanismo y estrategia para controlar o disminuir problemas relacionados con la salud como el sobrepeso y la obesidad[69,70]. También se ha postulado como uno de los elementos necesarios para aumentar la práctica de AF y mantenerla a lo largo de la vida[66]. No obstante, existe falta de acuerdo, estandarización y herramientas adecuadas para la evaluación de dicha CM.

Por esta razón, se propone este trabajo de investigación que trató de desarrollar una herramienta digital para evaluar de forma rápida y sencilla la competencia motriz de niños y niñas de entre 6 y 12 años. Previamente a la creación de esta herramienta se realizaron tres estudios (una revisión sistemática y dos observacionales) que trataron de examinar aspectos relacionados con la fiabilidad de la evaluación de las HMB, considerada como requisito imprescindible para realizar una evaluación rigurosa.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3 HIPÓTESIS

H.1 – Existe suficiente literatura científica sobre la fiabilidad del *Test of Gross Motor Development (TGMD)* para realizar una valoración crítica, un resumen de los resultados y una recomendación del test como instrumento robusto para la evaluación de las habilidades motrices básicas.

H.2 – La fiabilidad inter- e intra-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* entre evaluadores expertos es superior a la de los evaluadores novatos y mediante la evaluación a través de vídeo es superior a la evaluación en directo.

H.3 – La fiabilidad inter-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* mejora tras un acuerdo previo entre los evaluadores de los criterios de rendimiento de cada habilidad.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4 OBJETIVOS

O.1 – Identificar, sintetizar y evaluar los estudios de investigación existentes relacionados con la fiabilidad del *Test of Gross Motor Development (TGMD)* a través de una revisión sistemática de la literatura.

O.2 – Estudiar la fiabilidad inter- e intra-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* entre evaluadores expertos y novatos a través de observaciones en directo y en vídeo.

O.3 – Determinar la fiabilidad inter-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* tras un acuerdo previo de los evaluadores sobre los criterios de rendimiento de cada habilidad.

O.4 – Diseñar una herramienta informática, *Alfamov*, que permita la evaluación sistemática de las habilidades motrices básicas en sus tres grupos diferenciados: locomotrices, manipulativas y de estabilidad.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

5 METODOLOGÍA

La presente tesis doctoral sigue el formato de **tesis monográficas con reproducción de publicaciones científicas**. Esto implica que parte de los resultados de la tesis doctoral son artículos publicados en revistas científicas con un índice de impacto relativo. Concretamente, en esta tesis doctoral se presentan, como resultados principales en formato de artículos, dos publicaciones indexadas y otro artículo en proceso de revisión. Los artículos publicados, en su formato final de publicación se pueden visualizar en el ANEXO 1.

Siguiendo este criterio, se presentan en el correspondiente apartado de resultados:

a) Una revisión sistemática de la literatura sobre la fiabilidad del Test of Gross Motor Development (TGMD), uno de los instrumentos más utilizados para la evaluación de las HMB [artículo publicado en PloS ONE, segundo cuartil del Journal Citation Report (JCR)].
b) Un estudio observacional que analizó la fiabilidad inter- e intra-observador del TGMD-3 entre evaluadores expertos y novatos a través de dos métodos de evaluación (en directo y en vídeo) [artículo original publicado en International Journal of Environmental Research and Public Health, primer cuartil del JCR].
c) Un estudio observacional que examinó la fiabilidad inter-observador del TGMD-3 tras el consenso de los evaluadores sobre los criterios de rendimiento de las HMB previo a la evaluación [artículo sin publicar que se encuentra en proceso de revisión].

Cada una de las publicaciones mencionadas cuenta de manera individual con una introducción, objetivos, metodología, resultados, discusión y conclusiones específicas.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Además, se presenta como resultado de esta tesis doctoral el diseño de una herramienta informática para la evaluación de las habilidades motrices básicas (*véase apartado 6.2 de resultados*). También se muestran otras contribuciones de la doctoranda en formato de artículos, presentaciones en congresos, estancias de investigación y premios que han contribuido a su formación como investigadora.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6 RESULTADOS

6.1 RESULTADOS DERIVADOS DE LA TESIS DOCTORAL

6.1.1 Artículo 1. Fiabilidad del Test of Gross Motor Development: una revisión sistemática

Referencia de la publicación: Rey E, **Carballo-Fazanes A**, Varela-Casal C, Abelairas-Gómez C; on behalf of ALFA-MOV Project collaborators. Reliability of the test of gross motor development: A systematic review. **PLoS ONE. 2020 16;15(7): e0236070.**

Revista: PLoS ONE (Editorial: Public Librarby Science, ISSN: 1932-6203)

Ezequiel Rey[1] (primer autor)
Aida Carballo-Fazanes[2,3] (segunda autora)
Cristina Varela-Casal[1] (tercera autora)
Cristian Abelairas-Gómez[2,3,4,*] (último autor)
on behalf of ALFA-MOV Project collaborators*

[1] Facultad de Educación y Ciencias del Deporte, Universidade de Vigo, España.

[2] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, España.

[3] Grupo de Investigación CLINURSID, Departamento de Psiquiatría, Radiología, Salud Pública, Enfermería y Medicina, Universidade de Santiago de Compostela, España.

[4] Facultad de Ciencias de la Educación, Universidade de Santiago de Compostela, España.

**Colaboradores ALFA-MOV: Roberto Barcala Furelos, José Eugenio Rodríguez Fernández y Carmen Neiro Rey.*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen

Objetivo: Identificar, sintetizar y evaluar los estudios que investigaron la fiabilidad de las variantes del Test of Gross Motor Development (TGMD).

Métodos: Se realizó una revisión sistemática para identificar los estudios que han investigado la consistencia interna, la fiabilidad inter-observador, intra-observador y test-retest de las variantes del TGMD a través de las bases de datos Scopus, Pubmed/MEDLINE, PsycINFO, Sport Discus y Web of Science.

Resultados: De los 265 estudios identificados, se incluyeron 23. La consistencia interna fue evaluada en 14 estudios, se confirmó una consistencia de buena a excelente para la puntuación global y el cociente motor general (CMG), y niveles de aceptables a excelentes en ambas subescalas (locomotrices y habilidades de control de objetos). La fiabilidad inter-observador, evaluada en 19 estudios, mostró valores del coeficiente de correlación intraclase (CCI) de buenos a excelentes en la puntuación de las habilidades locomotrices, la puntuación de las habilidades de control de objetos, la puntuación general y el CMG. La fiabilidad intra-observador, evaluada en 13 estudios, mostró valores de CCI excelentes en la puntuación global y el CMG, y valores de CCI de buenos a excelentes en la puntuación de las habilidades locomotrices y la puntuación de las habilidades de control de objetos. La fiabilidad test-retest se evaluó en 15 estudios con el 100% de los datos reportados por encima del umbral de fiabilidad aceptable cuando no se utilizó el CCI. Los estudios que utilizaron el estadístico CCI mostraron valores de buenos a excelentes en la puntuación de las habilidades de control de objetos, la puntuación global y el CMG; y valores de moderados a excelentes en la puntuación de las habilidades locomotrices.

Conclusiones: En general, los resultados de esta revisión sistemática indican que, independientemente de la variante del test, el TGMD tiene una consistencia interna de moderada a excelente, una fiabilidad

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

inter-observador de buena a excelente, una fiabilidad intra-observador de buena a excelente y una fiabilidad test-retest de moderada a excelente. Teniendo en cuenta los pocos estudios de alta calidad sobre la consistencia interna, sería recomendable realizar más estudios en este campo para mejorar su calidad. Dado que no existe un estándar de referencia para evaluar las HMB, las variantes del TGMD podrían ser apropiadas cuando se opte por una prueba psicométrica robusta. Sin embargo, parece necesario que tanto los investigadores como los profesionales cuenten con protocolos de formación estandarizados para la codificación de las variantes del TGMD, con el fin de garantizar una fiabilidad aceptable.

INTRODUCCIÓN

Las habilidades motrices básicas (HMB) se consideran "*pilares de movimiento*" para la realización de movimientos más avanzados y complejos, esenciales para la participación en muchos juegos organizados y no organizados, deportes u otras actividades físicas específicas [1-3]. Las HMB se clasifican generalmente en habilidades locomotrices (p. ej.: la carrera y el salto), habilidades de manipulación o de control de objetos (p. ej.: la recepción y lanzamiento) y habilidades de estabilidad (p. ej.: el equilibrio y los giros) [1,4]. La evidencia reciente sugiere que la competencia en HMB se asocia con mejores resultados de salud en los niños y, además, puede tener un papel potencial en la promoción de trayectorias de salud positivas a lo largo de la vida [5]. Sin embargo, el dominio de las HMB no surge de forma natural [6], y la información aprendida y los factores ambientales parecen desempeñar un papel importante en la consecución de un nivel de competencia en el periodo comprendido entre la primera infancia (2-3 años) y la infancia posterior (7-10 años) [7].

En vista de los beneficios para la salud previamente reportados, los instrumentos utilizados para evaluar y monitorizar la competencia motriz han ganado relevancia en la educación física durante las últimas décadas con el fin de identificar a los estudiantes con

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

discapacidad motriz, describir los niveles de competencia motriz y apoyar las decisiones curriculares en las escuelas [8]. Las herramientas de evaluación de las HMB pueden clasificarse en dos categorías: pruebas orientadas a la cantidad/producto o pruebas orientadas a la calidad/proceso [4,9]. Las pruebas orientados al producto evalúan cuantitativamente el resultado del movimiento (es decir, a qué distancia o a qué altura) [10]. Por otro lado, las pruebas de evaluación orientadas al proceso evalúan la presencia o ausencia de patrones de movimiento mostrados por un niño o niña, proporcionando información cualitativa sobre la competencia motriz de los niños que puede utilizarse para diseñar y planificar intervenciones [9,11]. Entre las herramientas de evaluación orientadas al proceso, el Test of Gross Motor Development (TGMD) y sus variantes (Test of Gross Motor Development -Segunda Edición [TGMD-2] y Test of Gross Motor Development -Tercera Edición [TGMD-3]) son, probablemente, los test más utilizados para evaluar la competencia de las HMB en entornos educativos, clínicos y de investigación debido a su bajo coste y viabilidad [12-15]. El TGMD es una herramienta de evaluación normativa y basada en criterios, diseñada para evaluar cualitativamente el rendimiento de las habilidades motrices gruesas de niños de entre 3 y 10 años y 11 meses, con y sin discapacidades [13-15].

El TGMD se compone de dos subescalas, locomotriz y de control de objetos, que evalúan de seis a siete HMB con entre tres y cinco criterios de rendimiento, dependiendo de la habilidad [14,15] (Tabla 1). El rendimiento del niño se puntúa con 1 o 0 dependiendo de la presencia o ausencia de dichos criterios y las puntuaciones brutas finales pueden convertirse en rangos percentiles y puntuaciones estándar. Los resultados del test pueden utilizarse para identificar a los niños con retraso en el desarrollo de la motricidad gruesa [16], para diseñar, planificar y evaluar el éxito de programas de intervención en el desarrollo de las HMB, para evaluar el progreso individual y para servir como herramienta de evaluación en la investigación [14].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 1[1]. TGMD-2 y TGMD-3 subescalas, habilidades, criterios de rendimiento, y puntuaciones.

TGMD-2						TGMD-3					
Locomotrices			**Control de objetos**			**Locomotrices**			**Control de objetos**		
Habilidades	NC	PM	Habilidades	NC	PM	Habilidades	NC	PM	Habilidades	NC	PM
Carrera	4	8	Bateo	5	10	Carrera	4	8	Bateo	5	10
Galope	4	8	Bote estático	4	8	Galope	4	8	Golpeo con raqueta	4	8
Salto a un pie	5	10	Recepción	3	6	Salto a un pie	4	8	Bote estático	3	6
Zancada	3	6	Patada	4	8	Skip	3	6	Recepción	3	6
Salto horizontal	4	8	Lanz. por encima	4	8	Salto horizontal	4	8	Lanz. por encima	4	8
Desplazamiento lateral	4	8	Lanz. rodando	4	8	Desplazamiento lateral	4	8	Lanz. por debajo	4	8
									Patada	4	8

NC número de criterios de rendimiento, *PM* puntuación máxima, *Lanz.* lanzamiento

[1]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

La fiabilidad puede considerarse un requisito previo a la aplicación clínica, educativa y de investigación de cualquier instrumento, más aún para las medidas de campo, como el TGMD. En este sentido, en los últimos años se han publicado varios estudios que han examinado la fiabilidad inter-observador, intra-observador y test-retest del TGMD en diferentes grupos de población, incluyendo niños con trastorno del espectro autista (TEA) [17], con trastorno por déficit de atención e hiperactividad [18], con deficiencias visuales [19], con trastornos mentales y conductuales [20] y con discapacidad intelectual [21]. Dada la creciente evidencia científica sobre este tema y la amplia aplicación de esta herramienta de evaluación, parece que está justificada una revisión sistemática de la fiabilidad del TGMD. Por lo tanto, este estudio tuvo como objetivo identificar, sintetizar y evaluar los estudios que investigaron la fiabilidad del TGMD y valorar críticamente y resumir sus resultados. Los resultados obtenidos pueden ayudar a clarificar la verdadera fiabilidad de esta prueba y, por tanto, proporcionar un recurso valioso para los profesionales e investigadores interesados en utilizar el TGMD o interpretar sus resultados.

METODOLOGÍA

Estrategia de búsqueda

Esta revisión sistemática exhaustiva se llevó a cabo de acuerdo con las guías *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) [22]. Las búsquedas fueron una combinación de términos MeSH y palabras libres organizadas en tres bloques: términos relacionados con el desarrollo motor, TGMD y fiabilidad (Archivo S1). Se realizaron a través de las siguientes bases de datos: Scopus, Pubmed/MEDLINE, PsycINFO, Sport Discus y Web of Science. Nuestra pregunta PICO (participantes, intervenciones, comparaciones, resultados) [23] fue la siguiente: ¿Es el TGMD una batería fiable (O) en términos de consistencia interna, fiabilidad inter-observador, intra-observador y test-retest (C) para evaluar las HMB

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(I) de preescolares y escolares (P)? La búsqueda se realizó el 08 de diciembre de 2019.

Criterios de inclusión y exclusión

Los criterios de inclusión se establecieron en función de nuestra pregunta PICO:

Participantes: se incluyeron los estudios con participantes en etapa preescolar (≥3 y <6 años) y/o escolar (≥6 y <12 años). Si los datos sobre la edad divergían entre los participantes, sólo se seleccionaron los estudios con ≤25% de la muestra fuera del rango de 3-11 años. Si sólo se informaba de los datos estadísticos de la edad, se seleccionaron aquellos con una media de edad dentro del rango 3-11 y con una media de edad + desviación estándar ≤12. Los estudios que omitieron datos sobre la edad se consideraron no elegibles.

Intervención: se incluyeron artículos en los que se evaluaron las HMB de preescolares y/o escolares con el TGMD o cualquier versión modificada (TGMD-2 / TGMD-3). Sólo se incluyeron los estudios de investigación que abarcaban todas las HMB del TGMD o todas las habilidades de una de las dos subescalas (locomotriz / habilidades de control de objetos).

Comparación: se consideraron las investigaciones que estudiaron la consistencia interna, la fiabilidad inter-observador, intra-observador y test-retest del TGMD.

Resultado: los resultados principales fueron a) la consistencia interna y b) la fiabilidad inter-observador, intra-observador y test-retest de las habilidades locomotrices, de control de objetos, globales y del cociente motor general (CMG). Los resultados secundarios fueron la evaluación de la fiabilidad de cada habilidad.

Tipo de estudio: se incluyeron artículos originales publicados en inglés, español o portugués (de Portugal y Brasil). No se exigió un tamaño mínimo de la muestra.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Criterios de exclusión: Se excluyeron los estudios cuyo objetivo principal no era evaluar la consistencia interna, la fiabilidad inter-observador, intra-observador y test-retest del TGMD. De este modo, no se incluyeron aquellos manuscritos con resultados secundarios o adicionales relativos a la fiabilidad. También se excluyeron los artículos de opinión, los resúmenes, las cartas al editor, las revisiones sistemáticas y los metaanálisis.

Selección de estudios y extracción de datos

Tanto el cribado como la elegibilidad fueron realizados de forma independiente por dos autores (E.R y C.A-G) para minimizar el posible sesgo. Si había desacuerdos, se consultó a una tercera revisora (A.C-F) para tomar una decisión. Los datos fueron extraídos de forma independiente por dos revisores (A.C-F y C.A-G) basándose en los requisitos mínimos recomendados en los *Criterios de inclusión y exclusión* [24], y luego fueron verificados de forma cruzada.

Calidad metodológica

La calidad de los estudios se evaluó mediante la lista de comprobación COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments), para revisiones sistemáticas [25,26], que incluye 10 casillas con todos los estándares necesarios para evaluar la calidad de un estudio en función de diferentes propiedades específicas [25]. Las casillas 4 y 6 se utilizaron para evaluar la consistencia interna y la fiabilidad, respectivamente. La lista de comprobación COSMIN evalúa *los requisitos del diseño* (1 ítem para la consistencia interna y 3 ítems para la fiabilidad), *los métodos estadísticos* (1 ítem para ambas casillas) y la presencia o no de *otros defectos importantes en el diseño o los métodos estadísticos* (en ambas casillas). Según la lista de comprobación COSMIN, cada ítem de cada casilla se califica como *calidad muy buena*, *adecuada*, *dudosa* o *inadecuada* [26]. La calidad de cada casilla se corresponde con la calificación más baja de cualquier ítem de la casilla. La evaluación del riesgo de sesgo fue valorada por dos revisores (A.C-F

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y C.A-G) utilizando las herramientas disponibles en el sitio web de COSMIN (www.cosmin.nl). Si había desacuerdos y no había consenso tras la discusión, se consultó a un tercer revisor (E.R) para tomar una decisión.

Estadísticas de los manuscritos

Debido a la gran variedad de análisis estadísticos observados en los estudios incluidos, se han utilizado diferentes tipos de estadísticos de fiabilidad. La consistencia interna se evaluó mediante el alfa de Cronbach. De acuerdo con las directrices de tamaño del coeficiente alfa recomendadas por George y Malery [27], se utilizaron los siguientes valores para interpretar el alfa de Cronbach: $\alpha > 0.9$ - Excelente, $\alpha > 0.8$ - Bueno, $\alpha > 0.7$ - Aceptable, $\alpha > 0.6$ - Cuestionable, $\alpha > 0.5$ - Pobre, y $\alpha < 0.5$ - Inaceptable. Para la interpretación de la fiabilidad inter-observador, intra-observador y test-retest, una correlación de Pearson > 0.80 [28] o Coeficiente de Correlación Intraclase (CCI) > 0.70 o Kappa > 0.70 [25,28] se calificó como "fiabilidad adecuada". Teniendo en cuenta que el CCI fue el estadístico más utilizado en los estudios incluidos, para una clasificación más específica de la fiabilidad, se utilizó la siguiente clasificación del CCI: los CCI inferiores a 0.50, entre 0.50 y 0.75, entre 0.75 y 0.90, y superiores a 0.9 se clasificaron como fiabilidad pobre, moderada, buena y excelente, respectivamente [29]. Por último, para el análisis de la fiabilidad de cada habilidad, se definió operativamente "fiabilidad adecuada" como ≥ 0.6 para los CCI, considerado como el nivel mínimo útil de acuerdo [30], suficiente para observar el movimiento humano con fines de cribado [31].

RESULTADOS

Resumen de los estudios

La búsqueda inicial recuperó 238 resúmenes y se identificaron 27 estudios adicionales a través de otros recursos (es decir, comprobando el listado de referencias de artículos) (Fig. 1). Se seleccionaron 142 resúmenes tras eliminar los duplicados y finalmente se incluyeron 23

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

estudios. Hubo un mayor uso del TGMD-2 [15 (65.2%)] frente al TGMD-3 [8 (34.8%)]. Diecinueve estudios analizaron la puntuación de las habilidades locomotrices y de control de objetos y la puntuación global o el CMG, al menos, en una de las medidas de fiabilidad (Tabla 2). Tres estudios sólo analizaron la puntuación de habilidades locomotrices y de control de objetos [32-34] y uno sólo las habilidades de control de objetos [35]. En la mayoría de los casos, se utilizó la grabación en vídeo para la evaluación (n = 19). Los estudios se llevaron a cabo en 14 países diferentes con participantes de edades comprendidas entre los 4 y los 9 años, siendo alrededor del 40% niñas. El tamaño de las muestras oscilaba entre 10 y 2674 participantes. La tabla 3 muestra los datos extraídos de los artículos en relación con la consistencia interna, la fiabilidad inter-observador, intra-observador y test-retest.

Consistencia interna

La consistencia interna se evaluó en 14 estudios (8 con el TGMD-2 frente a 6 con el TGMD-3) (Tabla 3). Se calcularon los coeficientes alfa en diferentes grupos o participantes para la puntuación de las habilidades locomotrices (n = 18), la puntuación de las habilidades de control de objetos (n = 18), la puntuación global (n = 12) y el CMG (n = 3). Diez estudios informaron de coeficientes alfa entre 0.7 y 0.9 para la puntuación de habilidades locomotrices (4 por encima de 0.9 [17,36-38]), 11 para la puntuación de habilidades de control de objetos (3 por encima de 0.9 [36-38]), 4 para la puntuación global (3 por encima de 0.9 [17,38,39]) y 3 para el CMG (2 por encima de 0.9 [21,36]).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

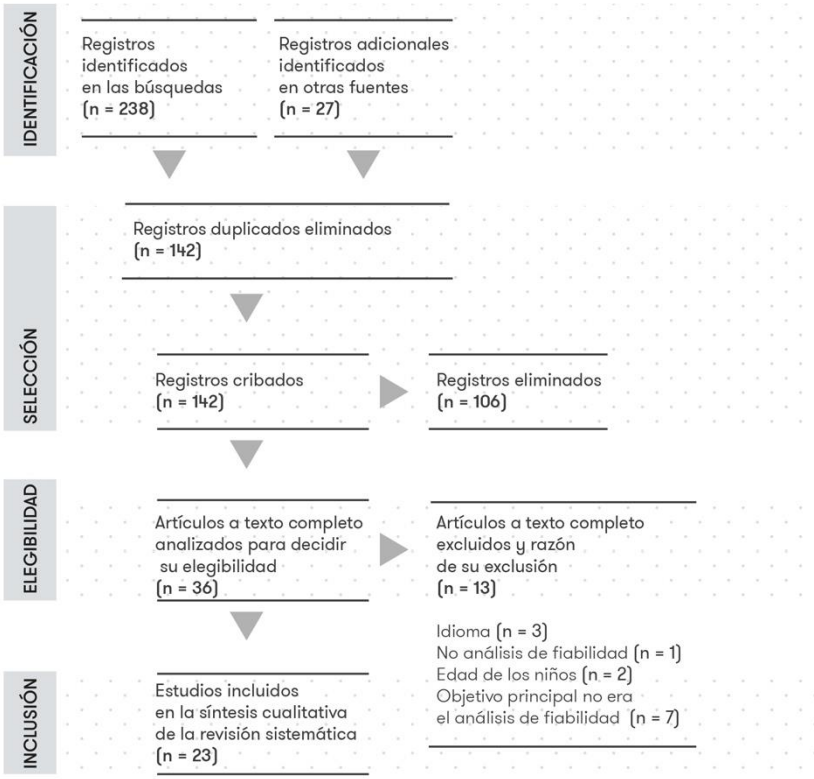


Figura 1[2]**.** Diagrama de flujo del proceso de búsqueda y selección de estudios.

[2]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 2[3]**.** Resumen de los estudios incluidos en la revisión sistemática.

Primer autor, año	Test	País	Muestra			Diseño	
			***N* y Perfil**	**Sexo**	**Edad (años)**	**Evaluación de fiabilidad**[a]	**Tipo de visualización**
Allen, 2017 [17]	TGMD-3 protocolo tradicional TGMD-3 protocolo de soporte visual	Australia	n=14 niños con TEA n=21 niños con desarrollo típico	Niños: 10 (71.4) Niñas: 4 (28.6) Niños: 12 (57.1) Niñas: 9 (42.9)	7.43 (2.03) 7.33 (1.75)	Consistencia Interna Fiabilidad inter-observador Fiabilidad intra-observador Fiabilidad test-retest	Vídeo
Ayán, 2019 [36]	TGMD-2	España	n=84 niños con desarrollo típico	Niños: 46 (54.8) Niñas: 38 (45.2)	8.35 (1.19)	Consistencia interna Fiabilidad inter-observador Fiabilidad test-retest	Vídeo
Aye, 2017 [59]	TGMD-2	Myanmar	n= 50 niños con desarrollo típico	Niños: 23 (46.0) Niñas: 27 (54.0)	5.40 (0.30)	Fiabilidad inter-observador Fiabilidad intra-observador(n=12) Fiabilidad test-retest (n=25)	Vídeo
Barnett, 2014 [35]	TGMD-2	Australia	n= 37 niños con desarrollo típico	Niños: 13 (35.0) Niñas: 24 (65.0)	6.20 (0.80)	Fiabilidad inter-observador	Directo
Cano-Cappellacci, 2015 [41]	TGMD-2	Chile	n=92 niños con desarrollo típico	Niños: 56 (60.9) Niñas: 36 (39.1)	7.50 (1.60)	Fiabilidad inter-observador (n=32) Fiabilidad intra-observador(n=38) Fiabilidad test-retest (n=32)	Vídeo

[3] Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

Primer autor, año	Test	País	Muestra			Diseño	
			N y Perfil	Sexo	Edad (años)	Evaluación de fiabilidad[a]	Tipo de visualización
Capio, 2016 [45]	TGMD-2	Filipinas	n=81 niños con discapacidad intelectual	Niños: 65 (80.2) Niñas: 16 (19.8)	9.29 (2.71)	Consistencia interna Fiabilidad inter-observador (n=10) Fiabilidad intra-observador(n=10)	Vídeo
Estevan, 2017 [47]	TGMD-3	España	n=178 niños con desarrollo típico	Niños 93 (52.5) Niñas 85 (47.5)	6.94 (1.89)	Consistencia interna Fiabilidad inter-observador (n=4) Fiabilidad intra-observador(n=4)	Vídeo
Farrokhi, 2014 [54]	TGMD-2	Irán	n= 1438 niños con desarrollo típico	Niños: 719 (50.0) Niñas: 719 (59.0)	6.53 (2.25)	Consistencia Interna Fiabilidad intra-observador(n=32) Fiabilidad test-retest (n=63)	Vídeo
Houwen, 2010 [19]	TGMD-2	Países Bajos	n=75 niños con discapacidad visual	Niños: 46 (61.0) Niñas: 29 (39.0)	8.50 (1.80)	Consistencia Interna Fiabilidad inter-observador (n=50) Fiabilidad intra-observador(n=25) Fiabilidad test-retest (n=23)	Vídeo
Kim, 2012 [43]	TGMD-2	Corea del Sur	n=22 niños con discapacidad intelectual	Niños: 16 (72.7) Niñas: 6 (27.3)	9.90 (1.30)	Fiabilidad inter-observador	Vídeo
Kim S, 2014 [57]	TGMD-2	Corea del Sur	n=141 niños con desarrollo típico	NI	6.80 (1.90)	Consistencia Interna Fiabilidad inter-observador (n=40) Fiabilidad test-retest (n=37)	Vídeo

Primer autor, año	Test	País	Muestra			Diseño	
			N y Perfil	Sexo	Edad (años)	Evaluación de fiabilidad[a]	Tipo de visualización
Kim C-I, 2014 [58]	TGMD-2	Corea del Sur	n=121 niños con desarrollo típico	Niños: 71 (58.7) Niñas: 50 (41.3)	5.98 (0.32)	Consistencia Interna	NI
Lopes, 2018 [40]	TGMD-2	Portugal	n=330 niños con desarrollo típico	Niños: 166 (50.3) Niñas: 164 (49.7)	7.90 (1.30)	Consistencia Interna Fiabilidad inter-observador Fiabilidad test-retest (n=22)	Vídeo
Maeng, 2017 [44]	TGMD-3	Estados Unidos	n=10 niños con desarrollo típico	Niños: 6 (60.0) Niñas: 4 (40.0)	6.57 (2.51)	Fiabilidad inter-observador Fiabilidad intra-observador	Vídeo
Mohammadi, 2019 [39]	TGMD-3	Irán	n=1600 niños con desarrollo típico	Niños: 800 (50.0) Niñas: 800 (50.0)	6.56 (2.29)	Consistencia Interna Fiabilidad inter-observador (n=160) Fiabilidad intra-observador(n=160) Fiabilidad test-retest (n=160)	Vídeo
Palmer, 2016 [33]	TGMD-2	Estados Unidos	n=43 niños con desarrollo típico	Niños: 25 (57.0) Niñas: 18 (43.0)	4.88 (0.28)	Fiabilidad inter-observador	Vídeo
Rintala, 2017 [42]	TGMD-3	Finlandia	n=60 niños con desarrollo típico	Niños: 28 (46.7) Niñas: 32 (53.3)	3-7[b]	Fiabilidad inter-observador (n=20) Fiabilidad intra-observador (n=20 evaluador A y n=20 evaluador B)	Vídeo
Simons, 2008 [21]	TGMD-2	Bélgica	n=99 niños con discapacidad intelectual	Niños: 67 (67.7) Niñas: 32 (32.3)	8.83 (1.75)	Consistencia Interna Fiabilidad inter-observador Fiabilidad test-retest (n=8)	Directo
Valentini, 2008 [32]	TGMD-2	Brasil	n=587 niños con desarrollo típico	Niños: 300 (51.1) Niñas: 287	7.52 (2.04)	Fiabilidad test-retest	Vídeo

Primer autor, año	Test	País	Muestra			Diseño	
			N y Perfil	Sexo	Edad (años)	Evaluación de fiabilidad[a]	Tipo de visualización
				(48.9)			
Valentini, 2012 [46]	TGMD-2	Brasil	n=2674 niños con desarrollo típico	Niños: 1352 (50.6) Niñas: 1322 (49.4)	7.56 (1.91)	Fiabilidad inter-observador Fiabilidad intra-observador Fiabilidad test-retest (n=648)	Vídeo
Valentini, 2017 [37]	TGMD-3	Brasil	n=597 niños con desarrollo típico	Niños: 295 (49.4) Niñas: 302 (50.6)	Niños: 6.76 (2.11) Niñas: 6.58 (2.06)	Consistencia Interna Fiabilidad inter-observador (n=50) Fiabilidad intra-observador(n=100) Fiabilidad test-retest (n=128)	Vídeo
Wagner, 2017 [34]	TGMD-3	Alemania	n=189 niños con desarrollo típico	Niños: 99 (52.4) Niñas: 90 (47.6)	7.15 (2.02)	Consistencia Interna Fiabilidad inter-observador (n=30) Fiabilidad intra-observador(n=30) Fiabilidad test-retest (n=104)	Evaluación en vídeo
Webster, 2017 [38]	TGMD-3	Estados Unidos	n=807 niños con desarrollo típico	Niños: 424 (52.5) Niñas: 338 (47.5)	6.33 (2.09)	Consistencia Interna Fiabilidad test-retest (n=30)	Evaluación en directo y en vídeo

[a]: Si la muestra utilizada en cada evaluación de la fiabilidad no se corresponde con el tamaño total de la muestra, se especifica entre paréntesis.
[b]: La edad se muestra como rango. No se informó de la media ni de la desviación estándar de todo el tamaño de la muestra.
TEA Trastorno del espectro autista, *CCI* Coeficiente de correlación intraclase, *NI* no informado.
Sexo expresado como frecuencias absolutas (frecuencias relativas); Edad expresada como media (desviación estándar).

Tabla 3[4]**.** Principales resultados

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
Allen, 2017 [17]	TGMD-3	TEA Protocolo de soporte visual(n=12): PHL: α=0.93; EEM=3.28 PHCO: α=0.81; EEM=4.39 Global: α=0.93; EEM=5.73 TEA protocolo tradicional(n=14): PHL: α=0.82; EEM=3.90 PHCO: α=0.75; EEM=4.63 Global: α=0.88; EEM=6.03 Desarrollo típico(n=21): PHL: α=0.70; EEM=3.20 PHCO: α=0.60; EEM=3.86 Global: α=0.74; EEM=5.18	TEA Protocolo de soporte visual(n=12): PHL: CCI=0.98, 95%IC (0.94-1.00); EEM=1.75 PHCO: CCI=0.96, 95%IC (0.86-0.99); EEM=2.02 Global: CCI=0.99, 95%IC (0.95-1.00); EEM=2.17 TEA protocolo tradicional(n=14): PHL: CCI=0.98, 95%IC (0.92-0.99) ; EEM=1.30 PHCO: CCI=0.97, 95%IC (0.91-0.99); EEM=1.60 Global: CCI=0.98, 95%IC (0.94-1.00);EEM=2.46 Desarrollo típico(n=21): PHL: CCI=0.91, 95%IC (0.79-0.96); EEM=1.76 PHCO: CCI=0.92, 95%IC (0.81-0.97); EEM=1.73 Global: CCI=0.94, 95%IC (0.87-0.98); EEM=2.49	TEA Protocolo de soporte visual(n=12): PHL: CCI=0.99, 95%IC (0.95-1.00); EEM=1.24 PHCO: CCI=1.00 95%IC (0.98-1.00); EEM=0.71 Global: CCI=0.99, 95%IC (0.98-1.00); EEM=2.17 TEA protocolo tradicional(n=14): PHL: CCI=0.97, 95%IC (0.88-0.99); EEM=1.59 PHCO: CCI=0.99, 95%IC (0.96-1.00); EEM=0.93 Global: CCI=0.99, 95%IC (0.92-1.00); EEM=1.74 Desarrollo típico(n=21): PHL: CCI=0.97, 95%IC (0.93-0.99); EEM=1.01 PHCO: CCI=0.91, 95%IC (0.68-0.97); EEM=1.83 Global: CCI=0.95, 95%IC (0.84-0.98); EEM=2.27	TEA Protocolo de soporte visual(n=8): PHL: CCI=0.92, 95%IC (0.62-0.98); EEM=2.99 PHCO: CCI=0.83, 95%IC (0.39-0.96); EEM=2.83 Global: CCI=0.92, 95%IC (0.66-0.98); EEM=4.62 TEA protocolo tradicional(n=8): PHL: CCI=0.92, 95%IC (0.65-0.98); EEM=2.24 PHCO: CCI=0.82 , 95%IC (0.31-0.96); EEM=2.93 Global: CCI=0.91 95%IC (0.63-0.98); EEM=4.06 Desarrollo típico(n=17): PHL: CCI=0.81, 95%IC (0.54-0.93); EEM=2.22 PHCO: CCI=0.84, 95%IC (0.62-0.94); EEM=2.45 Global: CCI=0.92, 95%IC (0.78-0.97); EEM=2.72

[4] Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
Ayán, 2019 [36]	TGMD-2	PHL: α=0.975 PHCO: α=0.963 CMG: α=0.974	PHL: CCI=0.976 PHCO: CCI=0.956 CMG: CCI=0.985	NI	PHL: r=0.952 PHCO: r=0.929 CMG: r=0.956
Aye, 2017 [59]	TGMD-2	NI	PHL: CCI=0.95 PHCO: CCI=0.88 CMG: CCI=0.89 Evaluador A *vs* Evaluador B: PHL: r=0.97 (p<0.001) PHCO: r=0.96 p<0.001) CMG: r=0.97 (p<0.001) Evaluador A *vs* Evaluador C: PHL: r=0.94 (p<0.001) PHCO: r=0.85 (p<0.001) CMG: r=0.87 (p<0.001) Evaluador B *vs* Evaluador C: PHL: r=0.93 (p<0.001) PHCO: r=0.96 p<0.001) CMG: r=0.88 (p<0.001)	PHL: CCI=0.98 PHCO: CCI=0.95 CMG: CCI=0.97	PHL: CCI=0.82 PHCO: CCI=0.79 CMG: CCI=0.76
Barnett, 2014 [35]	TGMD-2	NI	PHCO: CCI=0.93 95%IC (0.87-0.96)	NI	NI
Cano-Cappellaci, 2015 [41]	TGMD-2	NI	PHL: IVC=0.87, 95%IC (0.73-0.93) PHCO: IVC=0.88 (0.77-0.94) Global: IVC=0.86, 95%IC (0.72-0.93)	PHL: IVC=0.92, 95%IC (0.83-0.95) PHCO: IVC=0.86, 95%IC (0.76-0.93) Global: IVC=0.91, 95%IC (0.83-0.95)	PHL: IVC=0.86, 95%IC (0.71-0.93) PHCO: IVC=0.80, 95%IC (0.59-0.90) Global: IVC=0.88, 95%IC (0.75-0.94)

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
Capio, 2016 [45]	TGMD-2	PHL: α=0.830 PHCO: α=0.792	PHL: CCI=0.996, 95%IC (0.984-0.999) PHCO: CCI=0.998, 95%IC (0.992-1.000) Global: CCI=0.998, 95%IC (0.991-0.999)	PHL: CCI=0.995, 95%IC (0.978-0.999) PHCO: CCI=0.998, 95%IC (0.991-0.999) Global: CCI=0.997, 95%IC (0.989-0.999)	NI
Estevan, 2017 [47]	TGMD-3	PHL: α=0.80, 95%IC (0.75-0.84) PHCO: α=0.85, 95%IC (0.81-0.88) Global: α=0.89, 95%IC (0.87-0.92)	Global: CCI=0.90, 95%IC (0.66-0.98)	Global: CCI=0.98, 95%IC (0.85-1.00)	NI
Farrokhi, 2014 [54]	TGMD-2	PHL: α*=0.78 PHCO: α*=0.74 CMG: α*=0.80	NI	PHL: CCI=0.95, 95%IC (0.91-0.97) PHCO: CCI=0.99 95%IC (0.97-0.99) CMG: CCI=0.97, 95%IC (0.94-0.98)	PHL: CCI=0.65, 95%IC (0.50-0.79) PHCO: CCI=0.85, 95%IC (0.75-0.91) CMG: CCI=0.81, 95%IC (0.70-0.89)
Houwen, 2010 [19]	TGMD-2	PHL: α=0.71 PHCO: α=0.72	PHL: CCI=0.82, 95%IC (0.70-0.90) PHCO: CCI=0.93, 95%IC (0.88-0.96) Global: CCI=0.89, 95%IC (0.81-0.93)	PHL: CCI=0.85, 95%IC (0.69-0.93) PHCO: CCI=0.93, 95%IC (0.84-0.97) Global: CCI=0.95, 95%IC (0.88-0.98)	PHL: CCI=0.86, 95%IC (0.70-0.94) PHCO: CCI=0.87, 95%IC (0.72-0.94) Global: CCI=0.92, 95%IC (0.82-0.91)
Kim, 2012 [43]	TGMD-2	NI	PHL: CCI=0.80 PHCO: CCI=0.75 Global: CCI=0.78	NI	NI
Kim S, 2014 [57]	TGMD-2	PHL: α=0.82 PHCO: α=0.73	Evaluador A *vs* Evaluador B: PHL: CCI=0.94	NI	PHL: r=0.90 PHCO: r=0.85

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
		Global: α=0.87	PHCO: CCI=0.85 Global: CCI=0.97 Evaluador A *vs* Evaluador C: PHL: CCI=0.90 PHCO: CCI=0.80 Global: CCI=0.93 Evaluador B *vs* Evaluador C: PHL: CCI=0.91 PHCO: CCI=0.77 Global: CCI=0.92 Evaluador A *vs* Evaluador B *vs* Evaluador C PHL: CCI=0.94 PHCO: r=0.92 Global: CCI=0.96		
Kim C-I, 2014 [58]	TGMD-2	Pre-pesaje: PHL: α=0.53 PHCO: α=0.68 Global: α=0.72 Post-pesaje PHL: α=0.66 PHCO: α=0.53 Global: α=0.70	NI	NI	NI
Lopes, 2018 [40]	TGMD-2	PHL: α=0.46 PHCO: α=0.64 Global: α=0.69	*Kappa* La diferencia entre la medida del test y la del retest varió entre k=0,7 (consistencia	NI	*The Bland-Altman method* PHL: 95% del acuerdo varió 0.85-1.17

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
			moderada) y k=1 (consistencia perfecta).		Tasa de acuerdo: 1 (0.08) PHCO: 95% del acuerdo varió 0.63-1.16 Tasa de acuerdo: 0.80 (0.13) Global: 95% del acuerdo varió 0.80-1.13 Tasa de acuerdo: 0.96 (0.09)
Maeng, 2017 [44]	TGMD-3	NI	PHL: CCI=0.92, 95%IC (0.82-0.98) PHCO: CCI=0.93, 95%IC (0.84-0.98) Global: CCI=0.96, 95%IC (0.91-0.99)	Evaluador A: PHL: CCI=0.99, 95%IC (0.98-0.99) PHCO: CCI=0.97, 95%IC (0.90-0.99) Global: CCI=0.99, 95%IC (0.95-0.99) Evaluador B: PHL: CCI=0.97, 95%IC (0.90-0.99) PHCO: CCI=0.98, 95%IC (0.91-0.99) Global: CCI=0.98, 95%IC (0.91-0.99) Evaluador C: PHL: CCI=0.99, 95%IC (0.95-0.99) PHCO: CCI=0.97, 95%IC (0.87-0.99) Global: CCI=0.98, 95%IC	NI

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
				(0.93-0.99) Evaluador D: PHL: CCI=0.94, 95%IC (0.76-0.98) PHCO: CCI=0.93, 95%IC (0.73-0.98) Global: CCI=0.95, 95%IC (0.81-0.99) Evaluador E: PHL: CCI=0.99, 95%IC (0.98-0.99) PHCO: CCI=0.99, 95%IC (0.98-0.99) Global: CCI=0.99, 95%IC (0.98-0.99) Todos los Evaluadores: PHL: CCI=0.98, 95%IC (0.96-0.99) PHCO: CCI=0.96, 95%IC (0.94-0.98) Global: CCI=0.98, 95%IC (0.96-0.99)	
Mohammadi, 2019 [39]	TGMD-3	PHL: α=0.85 PHCO: α=0.85 Global: α=0.91	PHL: CCI=0.97, 95%IC (0.96-0.98) PHCO: CCI=0.98, 95%IC (0.97-0.98) Global: CCI=0.98, 95%IC (0.97-0.98)	PHL: CCI=0.98, 95%IC (0.98-0.99) PHCO: CCI=0.99, 95%IC (0.993-0.996) Global: CCI=0.99, 95%IC (0.994-0.997)	PHL: r=0.92 PHCO: r=0.94 Global: r=0.95

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
Palmer, 2016 [33]	TGMD-2	NI	Novice vs expert coders PHL: k=-0.001 PHCO: k=-0.004	NI	NI
Rintala, 2017 [42]	TGMD-3	NI	PHL: CCI=0.56 PHCO: CCI=0.64 Global: CCI=0.62	Evaluador A: PHL: CCI=0.69 PHCO: CCI=0.77 Global: CCI=0.75 Evaluador B: PHL: CCI=0.73 PHCO: CCI=0.73 Global: CCI=0.73	NI
Simons, 2008 [21]	TGMD-2	PHL: α=0.82 PHCO: α=0.86 CMG: α=0.90	PHL: r=1.00 PHCO: r=1.00 CMG: r=1.00	NI	PHL: r_s=0.90 PHCO: r_s=0.92 CMG: r_s=0.98
Valentini, 2008 [32]	TGMD-2	NI	NI	NI	PHL: r=0.82 PHCO: r=0.88
Valentini, 2012 [46]	TGMD-2	NI	PHL: CCI=0.88 PHCO: CCI=0.89	CCI=0.96	PHL: r=0.83 PHCO: r=0.91 Global: r=0.90
Valentini, 2017 [37]	TGMD-3	*Correlaciones entre las habilidades y las subpruebas:* PHL: α=0.63 PHCO: α=0.76 Global: α=0.74 *Correlaciones entre los criterios de rendimiento y las escalas y subescalas*	Evaluador 1 *vs* Evaluador 2 PHL: CCI=0.95 PHCO: CCI=0.97 Global: CCI=0.98	PHL: CCI=0.90 PHCO: CCI=0.85 Global: CCI=0.90	PHL: r=0.93 PHCO: r=0.81 Global: r=0.90

Primer autor, año	Test	Resultados de Fiabilidad			
		Consistencia interna	Fiabilidad inter-observador	Fiabilidad intra-observador	Test-Retest
		PHL: α=0.90 PHCO: α=0.88 Global: α=0.93			
Wagner, 2017 [34]	TGMD-3	PHL: α=0.76 PHCO: α=0.89	PHL: CCI=0.88, 95%IC (0.76-0.95) PHCO: CCI=0.97, 95%IC (0.94-0.99)	PHL: CCI=0.97, 95%IC (0.94-0.99) PHCO: CCI=0.99, 95%IC (0.98-1.00)	PHL: CCI=0.94, 95%IC (0.91-0.96) PHCO: CCI=0.98, 95%IC (0.97-0.99)
Webster, 2017 [38]	TGMD-3	PHL: α=0.92 PHCO: α=0.95 Global: α=0.96	NI	NI	PHL: CCI=0.97 PHCO: CCI=0.95 Global: CCI=0.97

TGMD Test of gross motor development, *PHL* Puntuación Habilidades Locomotrices, *PHCO* Puntuación Habilidades de Control de Objetos, *CMG* Coeficiente Motor General, *TEA* Trastorno del Espectro Autista, α Coeficiente Alpha de Cronbach, *EEM* Error Estándar de Medida, *Global* rendimiento global de la competencia motriz gruesa, *CCI* Coeficiente de Correlación intraclse, *IC* Intervalo de confianza, r Coeficiente de Correlación de Pearson, *NI* No informado, *IVC* Índice de validez de contenido, α* Alpha no especificado, *k* Kappa, r_s Correlación de Spearman.

Fiabilidad inter-observador

Se evaluó la fiabilidad inter-observador en 19 estudios (12 con el TGMD-2 frente a 7 con el TGMD-3). Se utilizaron 3 evaluadores en 5 estudios y 5 evaluadores en 1 estudio. El resto (13) utilizó 2 evaluadores. Se calculó el CCI en la mayoría de los estudios, excepto en 4 en los que se utilizó el coeficiente de correlación de Pearson [21], el estadístico Kappa [33,40] o el índice de validez de contenido (IVC) [41]. ≈70% de los estadísticos de fiabilidad inter-observador reportados con respecto a la puntuación de las habilidades locomotrices y de control de objetos, la puntuación global o el CMG fueron superiores a 0.9. Sólo dos estudios mostraron valores de fiabilidad inter-observador inferiores a 0.75 (CCI) [42] o 0.7 (Kappa) [33], el último comparando evaluadores expertos con novatos. Las puntuaciones de cada habilidad individual se reportaron en 6 estudios con más de ≈90% de los valores de fiabilidad inter-observador por encima de 0.6, ≈30% por encima de 0.9 (CCI calculado en todas las comparaciones) [35-37,42-44] (Figs. 2 y 3).

Fiabilidad intra-observador

Se evaluó la fiabilidad intra-observador en 13 estudios (6 con el TGMD-2 frente a 7 con el TGMD-3). Se utilizaron 2 evaluadores en 2 estudios [42,45] y 3 [46] y 5 [44] en 1 estudio. Un estudio no informó del número de evaluadores [47], y los estudios restantes (8) utilizaron 1 evaluador. El CCI y el IVC se utilizaron en 12 y 1 estudio respectivamente. El ≈85% de los estadísticos mostrados en relación con la puntuación de las habilidades locomotrices y de control de objetos, la puntuación global o el CMG fueron superiores a 0.9, >95% superiores a 0.75. Tres estudios informaron de datos específicos de las habilidades locomotrices y de control de objetos individualmente, con ≈90% de los valores de fiabilidad intra-observador por encima de 0.6, ≈35% por encima de 0.9 (CCI calculado en todas las comparaciones) [37,42,44] (Figs. 4 y 5).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

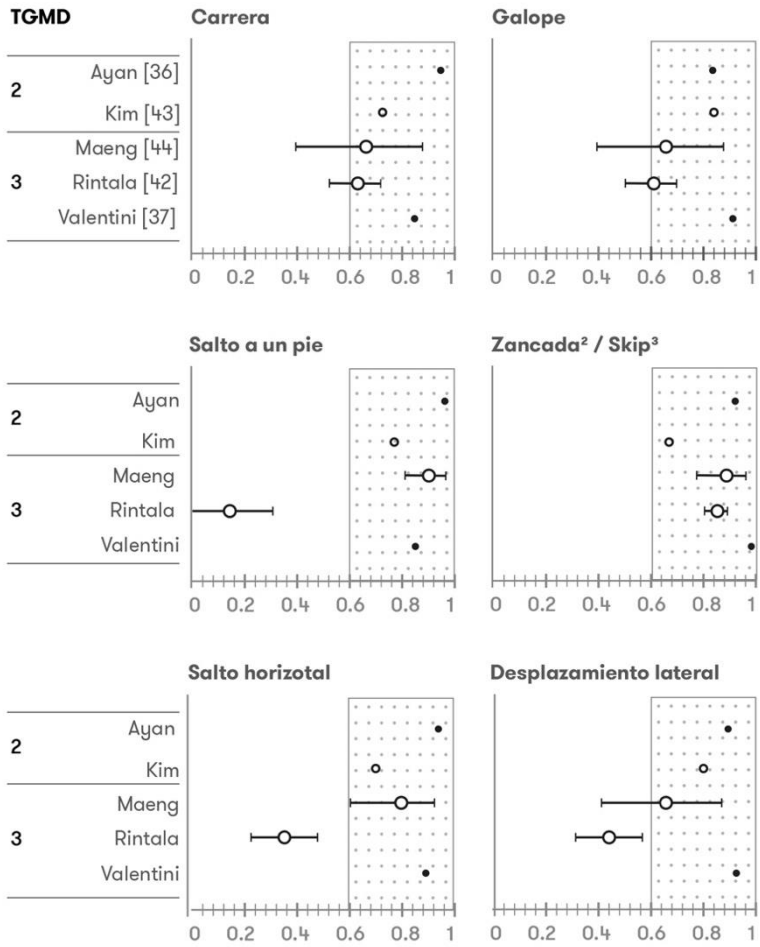


Figura 2[5]**.** Fiabilidad inter-observador de las habilidades locomotrices. *Círculos blancos*: evaluación de muy buena calidad según la lista de comprobación COSMIN; *Círculos negros*: evaluación de calidad adecuada según la lista de comprobación COSMIN; *Círculos más grandes*: CCI e IC 95%; *Círculos más pequeños*: CCI.

[5]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

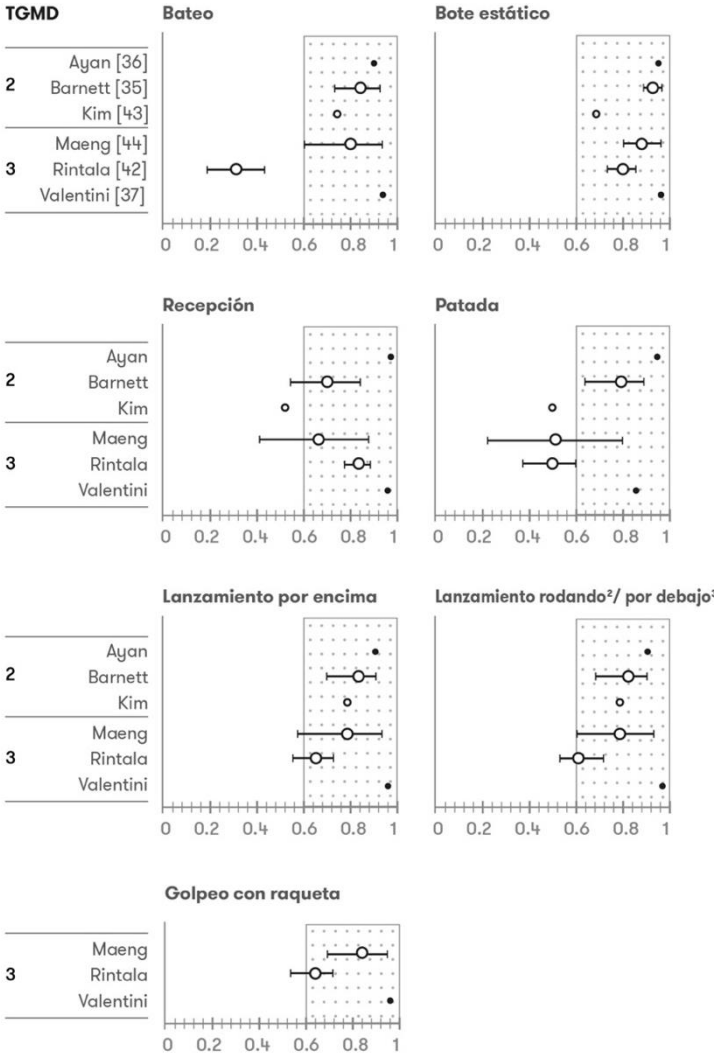


Figura 3[6]. Fiabilidad inter-observador de las habilidades de control de objetos. *Círculos blancos*: evaluación de muy buena calidad según la lista de comprobación COSMIN; *Círculos negros*: evaluación de calidad adecuada según la lista de comprobación COSMIN; *Círculos más grandes*: CCI e IC 95%; *Círculos más pequeños*: CCI.

[6]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

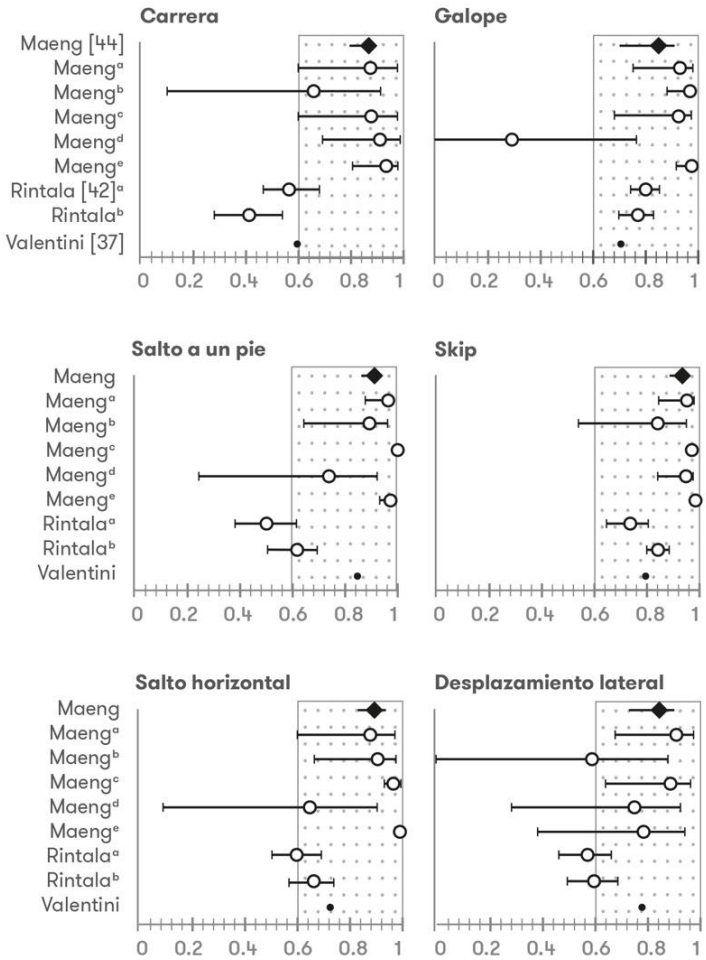


Figura 4[7]**.** Fiabilidad intra-observador de las habilidades locomotrices. *a-e*: Fiabilidad intra-observador de cada evaluador; *rombos:* fiabilidad intra-observador de todos los evaluadores; *Círculos blancos*: evaluación de muy buena calidad según la lista de comprobación COSMIN; *Círculos negros*: evaluación de calidad adecuada según la lista de comprobación COSMIN; *Círculos más grandes*: CCI e IC 95%; *Círculos más pequeños*: CCI.

[7]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

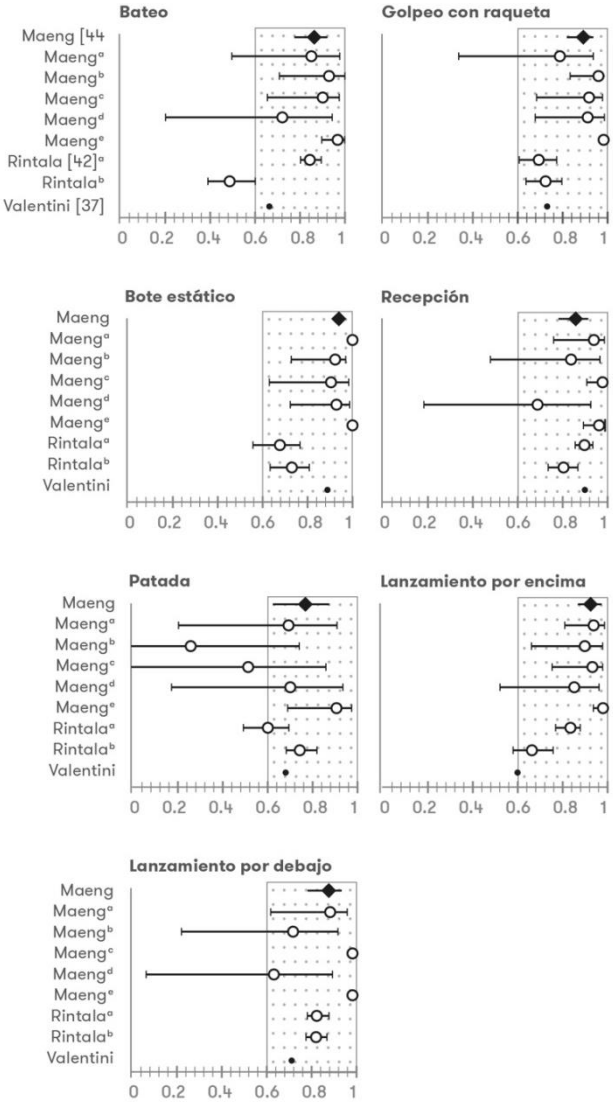


Figura 5[8]**.** Fiabilidad intra-observador de las habilidades de control de objetos. *[a-e]*: Fiabilidad intra-observador de cada evaluador; *rombos:* fianilidad intra-observador de todos los evaluadores; *círculos blancos*: evaluación de muy buena calidad según la lista de comprobación COSMIN; *Círculos negros*: evaluación de calidad adecuada según la lista de comprobación COSMIN; *Círculos más grandes*: CCI e IC 95%; *Círculos más pequeños*: CCI.

[8]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

Fiabilidad del test-retest

La fiabilidad test-retest se evaluó en 15 estudios (10 con el TGMD-2 frente a 5 con el TGMD-3). La fiabilidad test-retest se evaluó mediante el CCI (n=6), el coeficiente de correlación de Pearson (n=7), el CVI (n=1) y el ratio de acuerdo (n=1). La fiabilidad del TGMD medida a lo largo del tiempo mostró valores superiores a 0.8 en el 100% de las evaluaciones calculadas con el coeficiente de correlación de Pearson, el CVI y el ratio de acuerdo respecto a la puntuación de las habilidades locomotrices y de control de objetos, la puntuación global o el CMG. En cuanto al CCI, más del 95% de los valores fueron superiores a 0.75 y el 40% a 0.9. En tres estudios se calculó la fiabilidad test-retest para cada habilidad con valores de fiabilidad superiores a 0.8 en ≈50% (coeficiente de correlación de Pearson utilizado en todas las comparaciones) [36,37,46] (Fig. 6).

Niños con discapacidades

Cinco estudios analizaron alguna medida de fiabilidad en niños con discapacidad: TEA (TGMD-3) [17], discapacidad intelectual (TGMD-2) [21,43,45] y discapacidad visual (TGMD-2) [19]. Además, en el caso de los niños con TEA, se evaluaron ambos protocolos, el tradicional y el de apoyo visual. Se midió la consistencia interna en cuatro de los artículos con valores superiores a 0.7 en todas las mediciones relativas a las habilidades locomotrices y de control de objetos, la puntuación global y el CMG. Se evaluó la fiabilidad inter-observador en los cinco manuscritos, mientras que la fiabilidad intra-observador y la fiabilidad test-retest se comprobaron en tres artículos. Se observó una alta fiabilidad con puntuaciones superiores a 0.9 en ≈90% de los casos en términos de inter- e intra-observador; ≈70% en test-retest.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

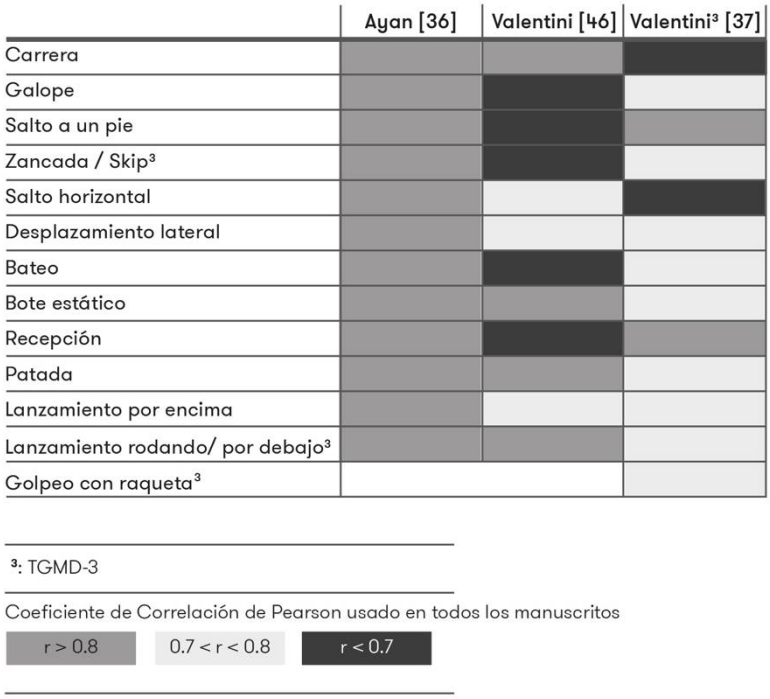


Figura 6[9]**.** Fiabilidad test-retest de las habilidades locomotrices y de control de objetos.

Calidad de los estudios

Un estudio se clasificó como *muy buena* calidad en términos de consistencia interna y otro como *insuficiente*. El resto de los artículos fueron clasificados como *dudosos*. El ítem de la lista de comprobación COSMIN con menor puntuación fue el referido al cálculo de los estadísticos de cada escala o subescala unidimensional por separado. La fiabilidad inter-observador se consideró *muy buena* en 8 estudios y *adecuada* en otros 8 (de 19 estudios). Se encontraron resultados

[9]Adaptada con permiso de Public Librarby Science, PloS ONE, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the test of gross motor development: A systematic review.* Rey et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

similares con respecto a la fiabilidad intra-observador, con 5 estudios con una evaluación *muy buena* y 5 con *adecuada* (de 13 estudios). La fiabilidad test-retest se clasificó como *adecuada* en 7 estudios (de 15 estudios). En la Tabla 4 se muestran los resultados más detallados de la evaluación de la calidad del COSMIN.

DISCUSIÓN

El TGMD es una de las diversas baterías de pruebas orientadas al proceso que pretenden evaluar la competencia motriz mediante la observación visual en niños de edad preescolar y primaria [8.48]. El propósito de esta revisión sistemática fue examinar la literatura relacionada con la fiabilidad del TGMD y realizar una valoración crítica y resumir sus resultados. En general, esta revisión reveló sólidas propiedades psicométricas tanto para el TGMD-2 como para el TGMD-3, lo que sugiere que las variantes del TGMD podrían ser una buena elección cuando se opta por un instrumento robusto en las pruebas de competencia motriz que utilizan enfoques orientados al proceso.

Consistencia interna

La consistencia interna se refiere al grado en que los componentes de la prueba (es decir, las habilidades en las variantes del TGMD) miden adecuadamente el mismo constructo (es decir, las subescalas y la puntuación total en las variantes del TGMD) [49]. Los resultados de los 14 estudios que evaluaron la fiabilidad mediante la consistencia interna confirmaron, en la mayoría de los casos, una consistencia de buena a excelente para la puntuación total del TGMD-2 y del TGMD-3 y el CMG, y niveles de consistencia interna de aceptables a excelentes en ambas subescalas (locomotriz y control de objetos), lo que indica que el instrumento parece ser consistente en la evaluación de las subescalas y la puntuación total [50]. Además, los criterios de rendimiento de las habilidades parecen abarcar una representación del mismo constructo [50].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fiabilidad inter-observador

La fiabilidad inter-observador muestra la concordancia o consistencia en las puntuaciones de dos o más evaluadores, y es una propiedad psicométrica esencial a la hora de evaluar la destreza del movimiento humano [35]. Los resultados de los 19 estudios que evaluaron la fiabilidad inter-observador confirmaron, en la mayoría de los casos, niveles adecuados de fiabilidad y valores de CCI de buenos a excelentes para el TGMD-2 y el TGMD-3 en la puntuación de las habilidades locomotrices, la puntuación de las habilidades de control de objetos, la puntuación global y el CMG. Se ha reportado ≈70% de los estadísticos por encima de 0.9 y el 100% de los valores de los coeficientes analizados por encima de los umbrales definitorios de la fiabilidad aceptable para la observación del cribado del movimiento humano. Sólo un estudio mostró niveles moderados de fiabilidad inter-observador para la puntuación de las habilidades locomotrices y de control de objetos y la puntuación global (en el TGMD-3) [42]; principalmente debido a la gran variabilidad observada entre tres habilidades individuales (salto a un pie, salto horizontal y golpeo con raqueta). Los valores de fiabilidad inter-observador encontrados en esta revisión sistemática fueron similares a los reportados utilizando otros instrumentos orientados al producto y al proceso como la *Movement Assessment Battery for Children-2nd edition* (MABC-2) [50], el *Bruininks-Oseretsky Test of Motor Proficiency–2nd Edition* (BOT-2) [51], *Basic Motor Competencies* (MOBAK) [52], o el *Dragon Challenge* [53].

Fiabilidad intra-observador

La fiabilidad intra-observador muestra el grado de acuerdo entre las evaluaciones repetidas de una prueba realizadas por el mismo evaluador. Esta revisión encontró excelentes valores de CCI de acuerdo intra-observador para las variantes del TGMD en la puntuación global y el CMG, y de buenos a excelentes en la puntuación de las habilidades locomotrices y la puntuación de las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

habilidades de control de objetos. Además, todos los estudios incluidos, excepto uno [42], informaron de niveles adecuados de fiabilidad intra-observador por encima de los umbrales de definición de fiabilidad aceptable para esta revisión sistemática [25,28].

Al igual que en el análisis de la fiabilidad inter-observador, sólo un estudio mostró niveles moderados de fiabilidad intra-observador para la puntuación de las habilidades locomotrices, la puntuación de las habilidades de control de objetos y la puntuación global del TGMD-3, principalmente debido a la gran variabilidad observada entre cinco habilidades individuales (carrera, bateo, desplazamiento lateral, salto con un pie y salto horizontal) [42]. En general, los resultados de la fiabilidad intra-observador de los estudios incluidos fueron algo más altos que los observados para la fiabilidad inter-observador. Esto apoya la evidencia de que es más probable que un evaluador esté de acuerdo de forma más consistente consigo mismo que con otros evaluadores [19], lo que se relaciona con la subjetividad y la discreción del evaluador [41]. Para minimizar la probabilidad de que un evaluador recuerde cómo ha puntuado el rendimiento de un niño concreto a partir de la puntuación anterior, el intervalo entre las evaluaciones se considera esencial. De hecho, el intervalo de tiempo utilizado en los estudios incluidos en esta revisión para reducir el sesgo influenciado por la memoria varía entre 12 días [54] y 3 meses [42]. Además, en tres de estos estudios no se ha especificado el intervalo [45-47]. Por lo tanto, los criterios relativos al intervalo de tiempo entre dos pruebas utilizados en los estudios de fiabilidad intra-observador analizados parecen deberse a una elección arbitraria. De esta forma, futuras investigaciones deberían comparar la fiabilidad intra-observador de las diferentes variantes del TGMD utilizando diferentes intervalos de tiempo para determinar el intervalo óptimo para minimizar el sesgo de memoria.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fiabilidad test-retest

La fiabilidad test-retest muestra la estabilidad temporal de las puntuaciones medidas por el mismo evaluador. Ambas variantes del TGMD revelaron niveles adecuados de fiabilidad test-retest, con el 100% de los estadísticos reportados por encima de los umbrales de definición de fiabilidad aceptable para esta revisión sistemática [25,28]. Específicamente, el TGMD-2 mostró valores de CCI de buenos a excelentes de fiabilidad test-retest (evaluados en 10 estudios) en la puntuación global, el CMG y la puntuación de habilidades de control de objetos, y valores de CCI de moderados a excelentes en la puntuación de habilidades locomotrices. El TGMD-3 (evaluado en 5 estudios) mostró excelentes valores de CCI de fiabilidad test-retest en el CMG, y valores de CCI de buenos a excelentes en las habilidades locomotrices y las habilidades de control de objetos. Los valores de fiabilidad test-retest observados en el TGMD-2 y en el TGMD-3 fueron similares a los reportados en otros instrumentos orientados al proceso que evalúan habilidades individuales de forma aislada, como *Victoria HMB Assessment* [55].

La familiarización de los participantes evaluados con los procedimientos del instrumento es un factor importante que puede influir en la fiabilidad de una prueba de rendimiento [56]. En este sentido, es importante señalar que los manuales del examinador del TGMD-2 y el TGMD-3 indican que cada participante debe completar sólo un ensayo de familiarización para cada habilidad tras la descripción verbal y la demostración práctica del evaluador [14,15]. Por lo tanto, basándonos en estos resultados, la fiabilidad de la prueba parece ser consistente independientemente de la variante del TGMD utilizada y del corto periodo de familiarización.

Adaptaciones culturales y lingüísticas

Las diferentes variantes del TGMD se utilizan de manera frecuente en varios países del mundo. Sin embargo, el TGMD se desarrolló para niños y niñas norteamericanos con un desarrollo típico. Debido a la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

relevancia sociocultural de las subescalas y de los criterios de rendimiento, varios estudios transculturales han investigado las propiedades psicométricas del TGMD-2 y del TGMD-3 en diferentes idiomas, como el español [36,41,47], el persa [39], el alemán [34] o el portugués [37,40,46] y/o culturas [57,58]. Las investigaciones realizadas al respecto han descrito unas características de fiabilidad elevadas y similares a las de la versión original, lo que evidencia la claridad de las instrucciones del TGMD y la falta de ambigüedad en la puntuación [39].

Evaluación en vídeo y en directo

Aunque el manual del examinador del TGMD no presupone la evaluación por vídeo [42], la mayoría de los estudios incluidos en esta revisión utilizaron evaluaciones a través de vídeo (n = 19). La evaluación mediante video del TGMD parece tener varias ventajas, ya que permite una puntuación más detallada, ayuda a la observación de los criterios de rendimiento difíciles con la repetición a cámara lenta, y hace posible reproducir cada movimiento tantas veces como sea necesario [42]. Además, consume menos tiempo en los entornos educativos, ya que la calificación de las pruebas puede realizarse fuera del horario de clase. Sin embargo, no siempre es posible realizar la evaluación mediante la grabación en vídeo del TGMD debido a diferentes consideraciones éticas o al equipo necesario [35]. En relación con esto, es importante señalar que los 3 estudios que analizaron la fiabilidad del TGMD utilizando la observación en directo mostraron excelentes valores de fiabilidad inter-observador [21,35], test-retest [21,38] y consistencia interna de buena a excelente [21,38]. La fiabilidad intra-observador no se evaluó utilizando la observación en directo en ninguno de los manuscritos incluidos en esta revisión sistemática. Por lo tanto, la fiabilidad de la variante del TGMD parece ser consistente independientemente del tipo de evaluación. Sin embargo, se necesitan más investigaciones para confirmar estos hallazgos, comparando la fiabilidad de las diferentes variantes del TGMD utilizando la evaluación en directo frente a la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

evaluación en vídeo, y la asociación entre la formación del evaluador y la capacidad de llevar a cabo la evaluación en directo.

Formación de evaluadores

Según el manual del examinador del TGMD, se recomienda la práctica supervisada en la administración e interpretación de las pruebas de desarrollo motor, de al menos tres evaluaciones previas antes de utilizar el TGMD en una situación real [14,15]. Sin embargo, la familiarización, la formación y la experiencia de los evaluadores en la administración del TGMD no se informó sistemáticamente ni se describió en los estudios incluidos en esta revisión sistemática. Además, la formación académica de los evaluadores es heterogénea, variando desde estudiantes de postgrado (Educación Física [35,36,42] y Ciencias del Deporte [36]), estudiantes de máster [57,58], estudiantes de doctorado [33,37,57], fisioterapeutas (*physical therapist*) [59], o fisioterapeutas pediátricos [45]. La evidencia previa resalta la necesidad de proporcionar protocolos de formación estandarizados para la codificación utilizando enfoques orientados al proceso como el TGMD-2 y el TGMD-3 para obtener resultados válidos y fiables [44]. Sin embargo, hasta donde sabemos, sólo un estudio analizó las diferencias de puntuación utilizando el TGMD-2 entre evaluadores expertos y novatos [33]. Los resultados mostraron que los evaluadores novatos (estudiantes de la Licenciatura de Educación Física con una sesión de formación de dos horas sobre el proceso de codificación) y los expertos (estudiante de doctorado en comportamiento motor con más de 3 años de experiencia en la codificación del TGMD-2) obtuvieron puntuaciones significativamente diferentes, excepto para la patada y el galope [33], lo que sugiere la necesidad de una formación más amplia hasta obtener un acuerdo. Por lo tanto, es necesario investigar en el futuro los efectos de proporcionar protocolos de formación estandarizados para la codificación de los datos del TGMD-2 y el TGMD-3 y determinar la formación mínima necesaria para garantizar niveles de fiabilidad aceptables. Además, investigaciones futuras deberían

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

examinar las subescalas y los criterios de rendimiento de cada habilidad en los que los evaluadores son más inconsistentes, para prestar especial atención durante la familiarización de los evaluadores.

Niños con discapacidades

Aunque la mayoría de los estudios incluidos en esta revisión han analizado la fiabilidad en niños con desarrollo típico, cinco estudios se realizaron con niños con discapacidad intelectual [21,43,45], con discapacidad visual [19] y con TEA [17]. En general, los valores de fiabilidad inter-observador (de bueno a excelente), intra-observador (de moderado a excelente) y test-retest (de bueno a excelente), observados fueron similares a los reportados en niños con desarrollo típico. En base a estos resultados, las variantes del TGMD podrían considerarse herramientas apropiadas para examinar las HMB en estas poblaciones. Sin embargo, el menor número de estudios realizados en niños con discapacidades abre una oportunidad para futuros estudios en estas y otras poblaciones especiales.

Fiabilidad de cada habilidad

La fiabilidad de cada habilidad del TGMD-2 y del TGMD-3 se evaluó en 7 estudios [35-37,42-44,46]. En general, se observaron niveles aceptables (CCI $\geq$ 0.6) de fiabilidad inter-observador para cuatro habilidades locomotrices (carrera, galope, zancada y skip) y cuatro habilidades de control de objetos (bote estático, lanzamiento por encima, lanzamiento por debajo rodando y golpeo con raqueta), mostrando valores de CCI de moderados a excelentes. Sin embargo, las tres habilidades locomotrices restantes (salto con un pie, salto horizontal y desplazamiento lateral) y las tres habilidades de control de objetos (bateo, recepción y patada) mostraron niveles contradictorios de fiabilidad entre los evaluadores. Las diferencias de fiabilidad entre las habilidades podrían ser un reflejo de la dificultad que entraña la evaluación de algunos componentes de las habilidades o de los criterios de rendimiento y de la necesidad de mejorar la claridad de su puntuación e interpretación.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

La fiabilidad intra-observador de las habilidades individuales fue algo mayor que la observada en la fiabilidad inter-observador, con siete habilidades (skip, salto horizontal, golpeo con raqueta, bote estático, recepción, lanzamiento por encima y lanzamiento por debajo), mostrando valores de CCI de moderados a excelentes. Las seis habilidades restantes (carrera, galope, salto con un pie, desplazamiento lateral, bateo y patada) mostraron niveles contradictorios de fiabilidad intra-observador, lo que puede reflejar la necesidad de un entrenamiento más intensivo en la evaluación de los criterios de rendimiento para estas habilidades específicas [44]. Es importante señalar que los tres estudios que analizaron la fiabilidad intra-observador de cada habilidad utilizaron la versión TGMD-3. Parece necesario realizar más investigaciones con el TGMD-2, que es la variante del test más utilizada en el contexto científico.

Tres estudios evaluaron la fiabilidad test-retest de cada habilidad individual del TGMD-2 [36,46] y del TGMD-3 [37]. Se encontraron varias discrepancias en este sentido en los estudios que utilizaron el TGMD-2. Ayán et al. [36] mostraron niveles aceptables de fiabilidad test-retest (coeficiente de correlación de Pearson ≥ 0.7) en todas las habilidades; sin embargo, Valentini [46] en siete habilidades (carrera, salto horizontal, desplazamiento lateral, bote estático, patada, lanzamiento por encima y lanzamiento por debajo rodando). En el caso del TGMD-3, sólo la carrera y el salto horizontal fueron las habilidades con valores bajos de fiabilidad test-retest, lo que podría reflejar mayores niveles de estabilidad temporal en el TGMD-3 que en el TGMD-2 [37]. Sin embargo, debido al escaso número de estudios, para profundizar en esta área puede ser necesario realizar futuras investigaciones que confirmen estos hallazgos.

Calidad metodológica

Catorce estudios evaluaron la consistencia interna, y sólo uno fue clasificado de *muy buena* calidad [36]. En ninguno de los estudios restantes se calcularon o expresaron los estadísticos para cada una de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

las escalas o subescalas unidimensionales como se ha destacado anteriormente [25]. Esto también implica el cálculo del alfa de Cronbach para cada habilidad. En cuanto a la fiabilidad inter-observador, intra-observador y test-retest, el ítem de métodos estadísticos fue el que más penalizó. Según la lista de control COSMIN, sólo utilizando el CCI (mostrando la fórmula o el modelo utilizado) para las puntuaciones continuas o el kappa para las puntuaciones dicotómicas/nominales, es posible conseguir una *muy buena* nota en este ítem [25]. Sin embargo, estudios anteriores sugieren que el coeficiente de variación podría utilizarse en este sentido con gran aplicabilidad [60,61]. Aun así, la mayor parte de los manuscritos en los que se evaluó la fiabilidad inter- e intra-observador fueron clasificados de *muy buena/adecuada* calidad.

El coeficiente de correlación de Pearson se utilizó en la mayoría de los manuscritos para evaluar la fiabilidad test-retest. Sin embargo, este estadístico no se considera el más adecuado para evaluar la fiabilidad [25,29,62]. Además, la evidencia de que los participantes se mantuvieran estables entre ambas evaluaciones es obligatoria para ser clasificada como *muy buena*. Dado que se podría considerar altamente probable que los niños se mantuvieran estables durante la evaluación, pero en varias ocasiones no se aportaron pruebas, la mayoría de los manuscritos se clasificaron en este apartado como *adecuados*. Debido a estos aspectos metodológicos rigurosos y exigentes en cuanto a los participantes y las estadísticas, ningún estudio fue clasificado como de *muy buena calidad*.

Limitaciones

Una primera limitación de esta revisión sistemática puede identificarse en los criterios específicos de elegibilidad que excluyen la llamada *literatura gris*. Por lo tanto, publicaciones relevantes podrían no haber sido incluidas en esta síntesis (es decir, monografías, resúmenes de conferencias y tesis). Además, sólo se seleccionaron las publicaciones en inglés, español o portugués que investigaban como

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

objetivo principal la fiabilidad del TGMD. Cabe suponer que algunos artículos significativos podrían haber sido publicados en otros idiomas. Otra limitación fue la ausencia de cualquier forma de meta-análisis debido a la amplia variedad de procedimientos estadísticos empleados para determinar la fiabilidad y la heterogeneidad de los participantes. Por último, dado que la variante TGMD-3 es una prueba relativamente nueva, el número de estudios incluidos que analizaron las propiedades psicométricas de esta versión fue significativamente inferior al del TGMD-2.

CONCLUSIONES

En esta revisión sistemática se consideraron un total de 23 estudios. En general, los resultados indican que, independientemente de la variante de la prueba y del tipo de evaluación (es decir, en directo o en vídeo), el TGMD tiene una consistencia interna de moderada a excelente, una fiabilidad inter-observador de buena a excelente, una fiabilidad intra-observador de buena a excelente y una fiabilidad test-retest de moderada a excelente. Además, la fiabilidad parece ser alta tanto en niños con desarrollo típico como en niños con discapacidades; sin embargo, el menor número de estudios en poblaciones especiales revela la necesidad de realizar más estudios de alta calidad. Dado que no existe un estándar de referencia para evaluar las HMB, las variantes del TGMD podrían ser apropiadas cuando se opta por una prueba psicométrica robusta. Sin embargo, los protocolos de entrenamiento estandarizados para la codificación de las variantes del TGMD parecen ser necesarios tanto para los investigadores como para los profesionales con el fin de garantizar una fiabilidad aceptable. No obstante, el protocolo de entrenamiento óptimo requiere más estudios. Por último, debido a los pocos estudios de alta calidad en términos de consistencia interna, sería recomendable que los futuros estudios en este campo utilizasen la lista de comprobación COSMIN para mejorar su calidad.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

REFERENCIAS

1. Gallahue D, Ozmun J, Goodway J. Understanding motor development: infants, children, adolescents, adults. 7th ed. New York, NY: McGraw-Hill; 2012.

2. Hulteen RM, Morgan PJ, Barnett LM, Stodden DF, Lubans DR. Development of Foundational Movement Skills: A Conceptual Model for Physical Activity Across the Lifespan. Sport Med. 2018; 48: 1533–40.

3. Logan SW, Ross SM, Chee K, Stodden DF, Robinson LE. Fundamental motor skills: A systematic review of terminology. J Sports Sci. 2018; 36: 781–96.

4. Burton A, Miller D. Movement skill assessment. Champaing, IL: Human Kinetics; 1998.

5. Robinson LE, Stodden DF, Barnett LM, Lopes VP, Logan SW, Rodrigues LP, et al. Motor Competence and its Effect on Positive Developmental Trajectories of Health. Sport Med. 2015; 45: 1273–84.

6. Palmer KK, Chinn KM, Robinson LE. Using Achievement Goal Theory in Motor Skill Instruction: A Systematic Review. Sport Med. 2017; 47: 2569–83.

7. Tompsett C, Sanders R, Taylor C, Cobley S. Pedagogical Approaches to and Effects of Fundamental Movement Skill Interventions on Health Outcomes: A Systematic Review. Sport Med. 2017; 47: 1795–819.

8. Scheuer C, Herrmann C, Bund A. Motor tests for primary school aged children: A systematic review. J Sports Sci. 2019; 37: 1097–112.

9. Yun J, Shapiro DR. A Quantitative Approach to Movement Skill Assessment for Children with Mental Retardation. Adapt Phys Act Q. 2004; 21: 269–80.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10. Barnett LM, Lai SK, Veldman SLC, Hardy LL, Cliff DP, Morgan PJ, et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. Sport Med. 2016; 46: 1663–88.

11. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. J Sci Med Sport. 2019; 22: 311–8.

12. Klingberg B, Schranz N, Barnett LM, Booth V, Ferrar K. The feasibility of fundamental movement skill assessments for pre-school aged children. J Sports Sci. 2019; 37: 378–86.

13. Ulrich D. Test of Gross Motor Development. Austin, TX: Pro-Ed; 1985.

14. Ulrich DA. Test of Gross Motor Development (2nd ed.). Austin, TX: Pro-Ed; 2000.

15. Ulrich DA. Test of Gross Motor Development (3nd ed.). Austin, TX: Pro-Ed; 2016.

16. Brian A, Pennell A, Taunton S, Starrett A, Howard-Shaughnessy C, Goodway JD, et al. Motor Competence Levels and Developmental Delay in Early Childhood: A Multicenter Cross-Sectional Study Conducted in the USA. Sport Med. 2019; 49: 1609–18.

17. Allen KA, Bredero B, Van Damme T, Ulrich DA, Simons J. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. J Autism Dev Disord. 2017; 47: 813–33.

18. Pan C-Y, Tsai C-L, Chu C-H. Fundamental Movement Skills in Children Diagnosed with Autism Spectrum Disorders and Attention Deficit Hyperactivity Disorder. J Autism Dev Disord. 2009; 39: 1694–705.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

19. Houwen S, Hartman E, Jonker L, Visscher C. Reliability and Validity of the TGMD-2 in Primary-School-Age Children With Visual Impairments. Adapt Phys Act Q. 2010; 27: 143–59.

20. Magistro D, Piumatti G, Carlevaro F, Sherar LB, Esliger DW, Bardaglio G, et al. Measurement invariance of TGMD-3 in children with and without mental and behavioral disorders. Psychol Assess. 2018; 30: 1421–9.

21. Simons J, Daly D, Theodorou F, Caron C, Simons J, Andoniadou E. Validity and Reliability of the TGMD-2 in 7–10-Year-Old Flemish Children with Intellectual Disability. Adapt Phys Act Q. 2008; 25: 71–82.

22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339: b2535.

23. Richardson W, Wilson M, Nishikawa J, Hayward R. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995; 123: A12–3.

24. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019) [Internet]. Cochrane. 2019.

25. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018; 27: 1171–9.

26. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018; 27: 1147–57.

27. George D, Mallery P. SPSS for Windows step by step: A Simple Guide and Reference, 11.0 update. 4th ed. Boston: Allyn & Bacon; 2003.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

28. Chinapaw MJM, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB. Physical Activity Questionnaires for Youth: a systematic review of measuremen propierties. Sport Med. 2010; 40: 539–63.

29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016; 15: 155–63.

30. Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax. 1991; 46: 454–6.

31. Moran RW, Schneiders AG, Major KM, Sullivan SJ. How reliable are Functional Movement Screening scores? A systematic review of rater reliability. Br J Sports Med. 2016; 50: 527–36.

32. Valentini N, Barbosa M, Cini G, Pick R, Spessate B, Balbinotti M. Teste de Desenvolvimento Motor Grosso: Validade e consistência interna para uma população Gaúcha. Rev Bras Cineantropometria Desempenho Hum. 2008; 10: 399–404.

33. Palmer KK, Brian A. Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. J Mot Learn Dev. 2016; 4: 142–51.

34. Wagner MO, Webster EK, Ulrich DA. Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation): Results of a Pilot Study. J Mot Learn Dev. 2017; 5: 29–44.

35. Barnett LM, Minto C, Lander N, Hardy LL. Interrater reliability assessment using the Test of Gross Motor Development-2. J Sci Med Sport. 2014; 17: 667–70.

36. Ayán C, Cancela J, Sánchez-Lastra M, Carballo-Roales A, Domínguez-Meis F, Redondo-Gutiérrez L. Fiabilidad y Validez de la Batería TGMD-2 en Población Española. Rev Iberoam Diagnóstico y Evaluación. 2019; 50: 21–34.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

37. Valentini NC, Zanella LW, Webster EK. Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. J Mot Learn Dev. 2017; 5: 15–28.

38. Webster EK, Ulrich DA. Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition. J Mot Learn Dev. 2017; 5: 45–58.

39. Mohammadi F, Bahram A, Khalaji H, Ulrich DA, Ghadiri F. Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. J Mot Learn Dev. 2019; 7: 106–21.

40. Lopes VP, Saraiva L, Rodrigues LP. Reliability and construct validity of the test of gross motor development-2 in Portuguese children. Int J Sport Exerc Psychol. 2018; 16: 250–60.

41. Cano-Cappellacci M, Leyton FA, Carreño JD. Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica. 2015; 49: 97.

42. Rintala PO, Sääkslahti AK, Iivonen S. Reliability Assessment of Scores From Video-Recorded TGMD-3 Performances. J Mot Learn Dev. 2017; 5: 59–68.

43. Kim Y, Park I, Kang M. Examining Rater Effects of the TGMD-2 on Children With Intellectual Disability. Adapt Phys Act Q. 2012; 29: 346–65.

44. Maeng H, Webster EK, Pitchford EA, Ulrich DA. Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition Among Experienced TGMD-2 Raters. Adapt Phys Act Q. 2017; 34: 442–55.

45. Capio CM, Eguia KF, Simons J. Test of gross motor development-2 for Filipino children with intellectual disability: validity and reliability. J Sports Sci. 2016; 34: 10–7.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

46. Valentini NC. Validity and Reliability of the TGMD-2 for Brazilian Children. J Mot Behav. 2012; 44: 275–80.

47. Estevan I, Molina-García J, Queralt A, Álvarez O, Castillo I, Barnett L. Validity and Reliability of the Spanish Version of the Test of Gross Motor Development–3. J Mot Learn Dev. 2017; 5: 69–81.

48. Cools W, Martelaer K De, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med. 2009; 8: 154–68.

49. Thomas J, Nelson J. Research methods in physical activity. 2nd ed. Champaign, IL: Human Kinetics; 1990.

50. Valentini NC, Ramalho MH, Oliveira MA. Movement Assessment Battery for Children-2: Translation, reliability, and validity for Brazilian children. Res Dev Disabil. 2014; 35: 733–40.

51. Bruininks R, Bruininks B. Bruininks-Oseretsky Test of Motor Proficiency–2nd Edition (BOT-2): Manual. Circle Pines, MN: AGS Publishing; 2005.

52. Scheuer C, Bund A, Becker W, Herrmann C. Development and validation of a survey instrument for detecting basic motor competencies in elementary school children. Cogent Educ. 2017; 4: 1337544.

53. Tyler R, Foweather L, Mackintosh KA, Stratton G. A Dynamic Assessment of Children's Physical Competence. Med Sci Sport Exerc. 2018; 50: 2474–87.

54. Farrokhi A, Zareh Z, Karimi A, Kazemnejad A, Ilbeigi S. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City. J Phys Educ Sport Manag. 2014; 5: 18–28.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

55. Lander N, Morgan PJ, Salmon J, Logan SW, Barnett LM. The reliability and validity of an authentic motor skill assessment tool for early adolescent girls in an Australian school setting. J Sci Med Sport. 2017; 20: 590–4.

56. Currell K, Jeukendrup AE. Validity, Reliability and Sensitivity of Measures of Sporting Performance. Sport Med. 2008; 38: 297–316.

57. Kim S, Kim MJ, Valentini NC, Clark JE. Validity and Reliability of the TGMD-2 for South Korean Children. J Mot Behav. 2014; 46: 351–6.

58. Kim C-I, Han D-W, Park I-H. Reliability and Validity of the Test of Gross Motor Development-II in Korean Preschool Children: Applying AHP. Res Dev Disabil. 2014; 35: 800–7.

59. Aye T, Oo KS, Khin MT, Kuramoto-Ahuja T, Maruyama H. Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. J Phys Ther Sci. 2017; 29: 1726–31.

60. Hopkins WG. Measures of Reliability in Sports Medicine and Science. Sport Med. 2000; 30: 1–15.

61. Grgic J, Oppici L, Mikulic P, Bangsbo J, Krustrup P, Pedisic Z. Test–Retest Reliability of the Yo-Yo Test: A Systematic Review. Sport Med. 2019; 49: 1547–57.

62. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60: 34–42.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.1.2 Artículo 2. Fiabilidad intra-observador (evaluación en directo vs vídeo) e inter-observador (evaluadores expertos vs novatos) del Test of Gross Motor Development – Third Edition

Referencia de la publicación: **Carballo-Fazanes A**, Rey E, Valentini NC, Rodríguez-Fernández JE, Varela-Casal C, Rico-Díaz J, Barcala-Furelos R, Abelairas-Gómez C. Intra-Rater (Live *vs* Video Assessment) and Inter-Rater (Expert *vs* Novice) Reliability of the Test of Gross Motor Development-Third Edition. **Int J Environ Res Public Health. 2021; 18(4):1652.**

Revista: International Journal of Environmental Research and Public Health. (Editorial: MDPI, ISSN: 1660-4601)

Aida Carballo-Fazanes[1,2] (primera autora)
Ezequiel Rey[3] (segundo autor)
Nadia C. Valentini[4] (tercera autora)
José Eugenio Rodríguez-Fernández[5] (cuarto autor)
Cristina Varela-Casal[3] (quinta autora)
Javier Rico-Díaz[5] (sexto autor)
Roberto Barcala-Furelos[1,3] (séptimo autor)
Cristian Abelairas-Gómez[1,2,5] (último autor)

[1] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, España.

[2] Grupo de Investigación CLINURSID, Departamento de Psiquiatría, Radiología, Salud Pública, Enfermería y Medicina, Universidade de Santiago de Compostela, España.

[3] Grupo de investigación REMOSS, Facultad de Educación y Ciencias del Deporte, Universidade de Vigo, España.

[4] Departmento de Educación Física, Fisioterapia y Danza, Universidade Federal do Rio Grande do Sul, Brazil.

[5] Facultad de Ciencias de la Educación, Universidade de Santiago de Compostela, España.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen: El Test of Gross Motor Development (TGMD) es una de las herramientas más comunes para evaluar las habilidades motrices básicas (HMB) en niños y niñas entre 3 y 10 años. El objetivo de este estudio fue examinar la fiabilidad intra e inter-observador del TGMD-Tercera edición (TGMD-3) entre evaluadores expertos y novatos utilizando la evaluación en directo y en vídeo. Cinco evaluadores [2 expertos y 3 novatos (uno de ellos licenciado en Educación Física y Ciencias del Deporte)] evaluaron y puntuaron la ejecución del TGMD-3 de 25 niños con desarrollo típico [Niñas: 60%; edad media (desviación estándar) 9.16 (1.31)]. Los escolares asistían a un colegio público de primaria de Santiago de Compostela (España) durante el curso 2019-2020. Los evaluadores puntuaron la ejecución de cada niño a través de dos modos de visualización (en directo y en vídeo). Se utilizó el Coeficiente de Correlación Intraclase (CCI) para determinar la concordancia entre los evaluadores. Nuestros resultados mostraron una fiabilidad intra-observador de moderada a excelente para la puntuación global y las subescalas de habilidades locomotrices y de control de objetos; una fiabilidad inter-observador de moderada a buena para la puntuación global y las habilidades de control de objetos; y de pobre a buena para la subescala locomotriz. Los evaluadores expertos y el evaluador novato con formación en educación física obtuvieron una mayor fiabilidad intra-observador que los evaluadores novatos. Sin embargo, la fiabilidad inter-observador fue más variable en todos los evaluadores, independientemente de su experiencia o formación. No se encontraron diferencias significativas en la fiabilidad al comparar las evaluaciones en directo y en vídeo. En la práctica clínica, sería recomendable que los evaluadores llegaran a un acuerdo antes de la evaluación para evitar interpretaciones subjetivas que pudieran distorsionar los resultados.

Palabras clave: TGMD-3; acuerdo; evaluación; habilidades motrices básicas; evaluadores; en directo; cámara lenta

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. Introducción

Las habilidades motrices básicas (HMB) consisten en movimientos básicos y organizados que implican la combinación de patrones de movimiento de dos o más partes del cuerpo [1]. Las HMB se consideran "*pilares del movimiento*" para movimientos más avanzados y complejos necesarios para participar en diferentes deportes, juegos y otras actividades físicas. Comúnmente, las HMB se clasifican en habilidades locomotrices (p. ej.: correr, saltar, deslizarse), habilidades de control de objetos (p. ej.: atrapar, patear, golpear, lanzar) y habilidades de equilibrio/estabilidad (p. ej.: equilibrio estático, equilibrio dinámico) [1-3].

Existen pruebas sólidas que sugieren asociaciones positivas entre la competencia en HMB y la actividad física [4], la condición física [4,5] y otros beneficios relacionados con la salud [6,7], como el estado de peso saludable [4,5], la aptitud cardiorrespiratoria [4,5] o los resultados académicos y cognitivos positivos [8], entre otros. Sin embargo, la adquisición de los niveles de competencia de las HMB no se produce de forma natural y es necesaria la implementación de programas estructurados de educación física en niños y niñas de edad preescolar y escolar [9].

En vista de los beneficios mencionados anteriormente, la literatura científica existente apoya la importancia de evaluar las HMB utilizando herramientas válidas y fiables [10], lo que también es necesario para proporcionar información valiosa sobre el rendimiento y el progreso motor de los niños [11]. Es esencial para la identificación temprana de posibles retrasos o trastornos que puedan afectar a la competencia motriz y al desarrollo cognitivo [12] y afectivo [13].

Existen varias herramientas de evaluación de las HMB, que pueden clasificarse a grandes rasgos en dos categorías: la evaluación orientada a la cantidad/producto, que evalúa el resultado del movimiento (p. ej.: la velocidad o la trayectoria) y la evaluación

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

orientada a la calidad/proceso, que evalúa los patrones de movimiento (p. ej.: cómo lanza una persona) [14,15]. Además, las evaluaciones pueden evaluar las habilidades motrices gruesas, movimientos que requieren el uso de grandes grupos motores (p. ej.: correr, saltar) [16] o las habilidades motrices finas, movimientos que implican pequeños grupos motores (p. ej.: escribir, comer) [17].

Una de las herramientas de evaluación más utilizadas para determinar la competencia de HMB en los niños y niñas es el Test of Gross Motor Development (TGMD) [3] y sus variantes TGMD-Segunda edición (TGMD-2) [18] y TGMD-Tercera edición (TGMD-3) [19]. El TGMD-3 es una herramienta de evaluación validada y fiable [20] orientada al proceso para evaluar la competencia motriz gruesa en niños de entre 3 años y 0 meses y 10 años y 11 meses. El TGMD-3 evalúa trece HMB divididas en dos subescalas, locomotriz y habilidades de control de objetos.

La fiabilidad es una característica fundamental a la hora de evaluar el rendimiento en investigación. Por un lado, la fiabilidad es una cuestión esencial, ya que permite a los investigadores replicar los estudios. Por otro lado, desde un punto de vista práctico, sería importante evaluar con precisión las variables que se quieren estudiar. Por ejemplo, en lo que respecta al ámbito del presente estudio, la evaluación de las HMB por parte de los docentes de la escuela o de los profesionales sanitarios. En este sentido, una revisión sistemática actual [20] de 23 estudios que evalúan la fiabilidad del TGMD mostró una fiabilidad inter-observador de buena a excelente, una fiabilidad intra-observador de buena a excelente y una fiabilidad test-retest de moderada a excelente. La mayoría de los estudios evaluaron la fiabilidad entre evaluadores con experiencia previa utilizando la evaluación mediante vídeo. Sin embargo, el TGMD es utilizado a veces por investigadores o profesionales (como profesorado de educación física) con poca o ninguna formación [21], ya que el manual del examinador sólo recomienda entrenar con al menos tres niños antes de una evaluación diagnóstica [18,19]. Por lo tanto, es

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

necesario conocer la fiabilidad entre evaluadores con diferente formación y experiencia. Sólo un estudio anterior ha comparado las diferencias de puntuación entre evaluadores expertos y novatos [22]. Los resultados indicaron que los evaluadores novatos no puntuaban el TGMD-2 de forma significativamente similar a los expertos [22]. Sin embargo, hasta la fecha, ningún estudio ha examinado esta cuestión en el TGMD-3.

Aunque el manual del examinador del TGMD no hace referencia a la evaluación de los niños mediante grabaciones de vídeo, muchos estudios utilizaron la evaluación grabada en vídeo debido a sus ventajas sobre la evaluación en directo. Permite evaluar los criterios de cada habilidad con más detalle y de forma repetida, incluso a cámara lenta si es necesario [22]. Todavía no se ha investigado la fiabilidad intra e inter-observador entre evaluadores expertos y novatos considerando un tipo diferente de visualización, en directo y a cámara lenta. Por lo tanto, el propósito de este estudio fue analizar la fiabilidad intra e inter-observador del TGMD-3 utilizando evaluaciones en directo y en vídeo entre evaluadores expertos y novatos.

2. Material y métodos

2.1. Participantes

Este estudio se llevó a cabo con 25 niños aparentemente sanos (60% niñas, n = 15; 100% hispanos, n = 25) de entre 8 y 10 años de edad (Media ± desviación típica (DT): 9.16 ± 1.31 años), que asistían a un colegio público de Educación Primaria durante el año académico 2019-2020 en Santiago de Compostela, Galicia, España. Se registraron mediciones antropométricas de cada niño (altura: 1.37 ± 0.11 m; peso: 34.57 ± 8.49 kg; e índice de masa corporal: 18.26 ± 2,95 kg-m^{-2}).

La participación en el estudio fue voluntaria y, previamente a su participación, los padres/madres o tutores legales firmaron un

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

consentimiento informado. Todos los niños recibieron información y dieron su consentimiento verbal antes de la prueba. Este estudio respetó los principios éticos de la Declaración de Helsinki y fue aprobado por el Comité Ético de la Universidade de Vigo, España.

2.2. Evaluadores

Cinco evaluadores (muestra de conveniencia), dos expertos (EE) y tres novatos (EN) se encargaron de evaluar las grabaciones en vídeo de los participantes. Los tres EN no tenían experiencia previa en la evaluación con el TGMD-3. Uno de los novatos era licenciado en Educación Física y Ciencias del Deporte (EN-EF); los otros dos eran un docente de primaria y una enfermera, ambos sin formación en educación física o ciencias del deporte ni experiencia en el trabajo con niños. Sin embargo, todos ellos podrían tener que evaluar la HMB de los niños en el futuro debido a su formación académica/profesional. Antes de evaluar la competencia de los niños, los EN estudiaron el contenido del manual del TGMD [19], preguntando a los EE sobre las dudas y practicando la administración del TGMD-3 a tres niños. Los dos EE eran licenciados en Educación Física y Ciencias del Deporte. Ambos con más de 5 años de experiencia utilizando el TGMD.

2.3. Test of Gross Motor Development - Third Edition [TGMD-3]

El TGMD-3 evalúa la competencia motriz gruesa de niños de entre 3 años y 0 meses y 10 años y 11 meses. Consta de dos subescalas: locomotriz y habilidades de control de objetos. La subescala locomotriz evalúa las habilidades motrices gruesas que implican una coordinación fluida del cuerpo mientras el niño se mueve en una u otra dirección e incluye: carrera, galope, skip, salto con un pie, salto horizontal y desplazamiento lateral. La subescala de habilidades de control de objetos evalúa las habilidades motrices gruesas que requieren eficacia en la interceptación y propulsión de objetos e incluye: bateo, golpe con raqueta, bote estático, recepción, patada, lanzamiento por encima y por debajo. Cada habilidad fue evaluada a través de tres a cinco criterios de rendimiento, puntuados con 1 o 0

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

según su presencia o ausencia [19]. Se obtuvieron puntuaciones independientes para cada habilidad, subescala y global a partir de las puntuaciones de los criterios.

Tras una descripción verbal y una demostración práctica, los niños realizan tres ensayos de cada habilidad. El primer ensayo es una práctica y no se puntúa. El evaluador puntúa los dos ensayos restantes de la siguiente manera: si el niño realiza correctamente un criterio de rendimiento, el evaluador puntúa con un "1" y si el niño no lo realiza correctamente, con un "0". Según esto, la puntuación máxima que puede alcanzar cada niño es de 100 puntos (46 para la subescala de locomotriz y 54 para la subescala de control de objetos).

2.4. Procedimientos

Un profesor de educación física y una enfermera (estudiante de doctorado) asistieron al colegio para realizarle las pruebas a los niños y niñas. Todos los datos se recogieron en marzo de 2020. Las pruebas se realizaron en un pabellón deportivo, en el horario habitual del colegio y siempre en presencia de al menos un docente del centro. Los evaluadores proporcionaron a los niños una descripción verbal de cada habilidad, seguida de un vídeo que mostraba la ejecución correcta de la habilidad (mostrado dos veces: a velocidad normal y a cámara lenta) siguiendo las directrices del TGMD-3 [19]. Cada niño realizó un ensayo de práctica seguido de dos intentos consecutivos, que fueron grabados en vídeo (cámara Nikon D5300) para un posterior análisis de fiabilidad. La administración de la prueba a cada niño duró aproximadamente 20-30 minutos.

Los cinco evaluadores evaluaron de forma independiente los vídeos grabados para analizar la fiabilidad inter- e intra-observador. Los observadores evaluaron el vídeo con dos modos de visualización: en directo (una vez, a velocidad normal y sin pausa) y a cámara lenta (viendo el vídeo tantas veces como fuera necesario, a cámara lenta y con pausas). Los niños, el orden de las habilidades y el tipo de visualización (en directo o a cámara lenta) de cada vídeo fueron

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

aleatorios para cada evaluador. Para el estudio de la fiabilidad intra-observador, cada evaluador volvió a evaluar los vídeos siguiendo el modo de visualización correspondiente según la aleatorización después de 2 semanas. En este sentido, se analizó la fiabilidad intra-observador según el modo de visionado (en directo *vs* a cámara lenta) y se estudió la fiabilidad inter-observador entre EN y EE en ambas condiciones, en directo y a cámara lenta.

2.5. Análisis estadístico

La fiabilidad se evaluó a través del coeficiente de correlación intraclase (CCI) tanto para la intra- como para la inter-observador. El proceso de selección del CCI se basó en el diagrama de flujo de Koo et al. [23]. Los valores del CCI y sus intervalos de confianza del 95% para la fiabilidad intra-observador se basaron en un modelo de efectos mixtos bidireccional con medidas únicas y acuerdo absoluto. En el caso de la fiabilidad entre evaluadores, medidas únicas, consistencia y modelo de efectos aleatorios bidireccional. Los valores inferiores a 0.5 indicaban una fiabilidad pobre, los valores entre 0.5 y 0.75 indicaban una fiabilidad moderada, los valores entre 0.75 y 0.9 indicaban una buena fiabilidad y los valores superiores a 0.90 indicaban una excelente fiabilidad.

Además, se realizó un ANOVA de medidas repetidas para analizar las puntuaciones globales y las de las subescalas locomotrices y de control de objetos; se proporcionaron las medias y las desviaciones estándar. Se integraron dos factores en el modelo: uno intragrupo (en directo *vs* a cámara lenta) y otro intergrupo (evaluadores). Los evaluadores se clasificaron de la siguiente manera: EN (Evaluador A y Evaluador B), EN-AF (Evaluador C) y EE (Evaluador D y Evaluador E).

Todos los análisis se realizaron con el paquete estadístico SPSS versión 23 (SPSS Inc, Chicago, IL). Se consideró un nivel de significación de $p < 0.05$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3. Resultados

3.1. Fiabilidad intra-observador: En directo vs cámara lenta

Los resultados de la fiabilidad intra-observador se muestran en la Tabla 1. Se encontró una fiabilidad de buena a excelente o excelente para la mayoría de las habilidades, principalmente en las evaluaciones de EE y EN-EF. Los valores del CCI de las puntuaciones de las subescalas y de las puntuaciones globales también mostraron una fiabilidad de buena a excelente o excelente en el caso de EE y EN-EF. Los EN mostraron una fiabilidad intra-observador de moderada a excelente, de buena a excelente o excelente, tanto en las puntuaciones de las subescalas como en las globales.

Los EE y el EN-EF alcanzaron una fiabilidad intra-observador al menos moderada en todas las habilidades (excepto un EE (Evaluador E): desplazamiento lateral). En el caso de los EN, la carrera (evaluador A y B), el salto con un pie (evaluador B), el salto horizontal (evaluador A y B), la patada (evaluador A) y el lanzamiento por debajo (evaluador B) fueron las habilidades en las que la fiabilidad fue de pobre a moderada o buena.

Los datos descriptivos relativos a la puntuación global y a las subescalas locomotrices y de control de objetos se presentan en la Tabla 2. En el análisis de medidas repetidas se encontraron diferencias significativas en el factor intragrupo (en directo *vs* a cámara lenta) en la subescala locomotriz (en directo: 33.7 ± 4.0; a cámara lenta: 33.2 ± 3.9; $p = 0.010$). El análisis post-hoc de Bonferroni reveló diferencias significativas entre las evaluaciones en directo y a cámara lenta en ambos EN (Evaluador A: $p = 0.014$/Evaluador B: $p = 0.036$) en la subescala locomotriz. Se encontraron las mismas diferencias en la subescala de control de objetos en uno de los EN (Evaluador B: $p = 0.001$) y en la puntuación global en el mismo EN (Evaluador B: $p < 0.001$) y en un EE (Evaluador D: $p = 0.038$).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 1[10]**.** Fiabilidad intra-observador*: en vivo *vs* a cámara lenta: Subescalas, habilidades y puntuaciones globales**.**

	EN		EN-EF	EE		Buena-a-excelente o Excelente
	Evaluador A	**Evaluador B**	**Evaluador C**	**Evaluador D**	**Evaluador E**	
Habilidades Locomotrices						**Número de evaluadores**
Carrera	0.304 (-0.409-0.676)	0.660 (0.236-0.850)	0.900 (0.773-0.956)	**0.913 (0.805-0.961)**	0.942 (0.869-0.975)	3/5
Desplazamiento lateral	0.808 (0.567-0.915)	0.921 (0.822-0.965)	**0.966 (0.924-0.985)**	-----†	0.714 (0.365-0.873)	2/4
Galope	0.950 (0.888-0.978)	0.925 (0.823-0.968)	**0.966 (0.923-0.985)**	**0.986 (0.969-0.994)**	**0.976 (0.944-0.989)**	5/5
Salto a un pie	0.825 (0.610-0.922)	0.705 (0.332-0.870)	0.936 (0.855-0.972)	**0.990 (0.977-0.996)**	0.884 (0.736-0.949)	2/5
Salto horizontal	0.765 (0.460-0.897)	0.680 (0.166-0.868)	**0.956 (0.900-0.981)**	**0.992 (0.982-0.997)**	0.916 (0.811-0.963)	3/5
Skip	0.892 (0.758-0.952)	0.861 (0.685-0.939)	**1.000**	**0.989 (0.976-0.995)**	**0.958 (0.905-0.981)**	4/5
Puntuación subescala locomotriz	0.865 (0.654-0.943)	0.817 (0.592-0.919)	**0.964 (0.918-0.984)**	**0.980 (0.954-0.991)**	0.901 (0.773-0.957)	3/5

[10]Adaptada con permiso de MDPI, International Journal of Environmental Research and Public Health, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the Test of Gross Motor Development-Third Edition.* Carballo-Fazanes et al. Copyright (2020).

	EN		EN-EF	EE		Buena-a-excelente o Excelente
	Evaluador A	**Evaluador B**	**Evaluador C**	**Evaluador D**	**Evaluador E**	
Habilidades de control de objetos						
Bateo	**0.959 (0.909-0.982)**	0.905 (0.783-0.958)	0.945 (0.876-0.976)	**0.968 (0.925-0.986)**	0.945 (0.876-0.976)	5/5
Bote estático	0.917 (0.813-0.964)	0.871 (0.662-0.947)	0.917 (0.803-0.964)	0.883 (0.739-0.948)	0.910 (0.798-0.960)	3/5
Lanzamiento por encima	0.931 (0.844-0.970)	0.887 (0.742-0.950)	0.934 (0.852-0.971)	**0.975 (0.942-0.989)**	0.915 (0.807-0.963)	4/5
Patada	0.659 (0.230-0.849)	0.846 (0.625-0.934)	0.839 (0.537-0.936)	0.948 (0.882-0.977)	**0.982 (0.956-0.992)**	2/5
Golpeo con raqueta	0.818 (0.582-0.920)	0.875 (0.720-0.945)	0.891 (0.745-0.953)	0.955 (0.898-0.980)	0.936 (0.856-0.972)	2/5
Recepción	**0.983 (0.961-0.992)**	0.899 (0.770-0.955)	0.908 (0.791-0.959)	**1.000**	**1.000**	5/5
Lanzamiento por debajo	0.916 (0.809-0.963)	0.698 (0.333-0.865)	**0.966 (0.923-0.985)**	**0.958 (0.901-0.982)**	0.821 (0.597-0.921)	3/5
Puntuación subescala control de objetos	**0.970 (0.934-0.987)**	0.879 (0.689-0.949)	0.949 (0.884-0.978)	**0.967 (0.915-0.986)**	**0.962 (0.915-0.983)**	4/5
Puntuación global	0.953 (0.895-0.979)	0.907 (0.716-0.963)	**0.968 (0.927-0.986)**	**0.982 (0.941-0.993)**	**0.961 (0.912-0.983)**	4/5

* CCI: medidas únicas, acuerdo absoluto, modelo de efectos mixtos bidireccional, intervalo de confianza del 95%; †: El evaluador D (EE) codificó la habilidad de desplazamiento lateral con las puntuaciones máximas en ambas pruebas (en vivo y en cámara lenta); en gris, habilidades con una fiabilidad intra-obsevador de buena a excelente (CCI y límite inferior ≥ 0,75); en gris y negrita, habilidades con una fiabilidad intra-evaluadora excelente (CCI y límite inferior ≥ 0,9). *EN*: Evaluador novato sin formación en educación física (A and B); *EN-EF*: Evaluador Novato con formación en Educación Física (C); *EE*: Evaluador Experto (D and E).

Tabla 2[11]**.** Descriptivos de la puntuación global y de las subescalas locomotriz y de control de objetos. Resultados expresados como media (desviación estándar).

Evaluadores		Locomotriz		Control de objetos		Global	
		Directo	C-L	Directo	C-L	Directo	C-L
EN	A	34.2 (3.4)	33.2 (3.1)	36.0 (5.3)	36.4 (4.9)	70.2 (7.7)	69.5 (6.3)
	B	34.8 (4.3)	33.9 (4.4)	35.2 (4.4)	33.8 (5.0)	70.0 (7.1)	67.6 (7.9)
EN-EF	C	32.5 (4.6)	32.3 (4.5)	33.9 (5.0)	34.8 (4.3)	66.4 (7.9)	67.2 (7.6)
EE	D	32.6 (4.0)	32.9 (4.0)	39.7 (5.6)	40.6 (5.5)	72.2 (8.6)	73.4 (8.4)
	E	34.5 (3.2)	33.8 (3.6)	37.6 (5.0)	37.6 (4.4)	72.2 (6.4)	71.5 (6.3)

EN: Evaluador novato sin formación en educación física (A and B); *EN-EF*: Evaluador Novato con formación en Educación Física (C); *EE*: Evaluador Experto (D and E); *C-L*: cámara lenta

3.2. Fiabilidad inter-observador: Novatos vs Expertos

En general, se encontró una fiabilidad inter-observador de pobre a buena tanto en las habilidades locomotrices como en las habilidades de control de objetos, con cambios nulos al comparar en directo *vs* a cámara lenta entre cada par de evaluadores. Las Tablas 3 y 4 presentan los resultados del CCI entre evaluadores para cada habilidad. La Figura 1 presenta las asociaciones del CCI con las subescalas locomotriz y de control de objetos y las puntuaciones globales.

En cuanto a las habilidades locomotrices, la fiabilidad fue de moderada a buena; en la mayoría de los casos, independiente de la experiencia o los antecedentes del evaluador. El galope fue la habilidad con la mejor fiabilidad entre los evaluadores; el salto horizontal y el skip alcanzaron valores de CCI superiores a 0.5 en la mayoría de las comparaciones. El salto con un pie fue la habilidad locomotriz con menor fiabilidad entre los evaluadores.

En cuanto a las habilidades de control de objetos, el bateo fue la habilidad con mayor fiabilidad entre los evaluadores, oscilando de

[11]Adaptada con permiso de MDPI, International Journal of Environmental Research and Public Health, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the Test of Gross Motor Development-Third Edition.* Carballo-Fazanes et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pobre a buena y de moderada a excelente. La recepción fue la habilidad con menor fiabilidad entre los evaluadores.

La subescala de habilidades de control de objetos y las puntuaciones globales fueron ligeramente superiores en EE en comparación con EN y EN-EF (Tabla 2). El análisis de medidas repetidas no mostró diferencias en las interacciones de los factores de modos de visualización y evaluadores para la subescala de habilidades locomotrices. Sin embargo, en lo que respecta a las habilidades de control de objetos, uno de los expertos (Evaluador D), codificó significativamente más alto en la evaluación en directo que un EN (Evaluador B: p = 0.024) y EN-EF (p = 0.001). Las mismas diferencias se encontraron en la evaluación a cámara lenta entre el evaluador D y ambos EN (evaluador A: p = 0.029/evaluador B: p = < 0.001) y EN-EF (p = 0.001). Por último, en cuanto las puntuaciones globales sólo se encontraron diferencias significativas entre un EE (Evaluador D) y EN-EF en la evaluación a cámara lenta (p = 0.031).

4. Discusión

El objetivo de este estudio fue analizar la fiabilidad intra- (en directo y a cámara lenta) e inter-observador [expertos y novatos (uno de ellos con formación en educación física)] del TGMD-3. En resumen, las subescalas de habilidades locomotriz y de control de objetos y las puntuaciones globales mostraron una fiabilidad intra-observador de moderada a excelente. La fiabilidad inter-observador mostró puntuaciones de moderadas a buenas para la subescala de habilidades de control de objetos y las puntuaciones globales tanto en directo como a cámara lenta, y una fiabilidad de pobre a buena para la subescala locomotriz también para ambos modos de evaluación.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 3[12]**.** Fiabilidad inter-observador* entre evaluadores. Habilidades locomotrices

Evaluadores	Test	Carrera	Desplazamiento lateral†	Galope	Salto con un pie	Salto horizontal	Skip	CCI ≥ 0.5
Evaluador AxB	Live	0.074 (−0.323-0.450)	0.739 (0.492-0.876)	**0.768 (0.541-0.891)**	0.525 (0.171-0.758)	0.624 (0.311-0.815)	**0.856 (0.701-0.934)**	5/6
	S-M	−0.002 (−0.390-0.386)	**0.815 (0.624-0.914)**	**0.747 (0.505-0.880)**	0.581 (0.249-0.791)	0.497 (0.135-0.742)	**0.747 (0.506-0.880)**	4/6
Evaluador AxC	Live	0.106 (−0.294-0.475)	**0.791 (0.581-0.902)**	0.610 (0.290-0.807)	0.477 (0.109-0.730)	0.586 (0.255-0.793)	**0.939 (0.866-0.973)**	4/6
	S-M	0.090 (−0.309-0.462)	0.437 (0.058-0.705)	0.687 (0.407-0.849)	0.251 (−0.152-0.583)	0.688 (0.409-0.849)	**0.805 (0.607-0.909)**	3/6
Evaluador AxD	Live	0.129 (−0.273-0.493)	-----	**0.801 (0.600-0.907)**	0.496 (0.134-0.742)	0.626 (0.314-0.816)	0.612 (0.294-0.808)	3/5
	S-M	−0.052 (−0.431-0.643)	-----	**0.893 (0.773-0.952)**	0.385 (−0.004-0.672)	0.598 (0.274-0.801)	0.588 (0.259-0.795)	3/5
Evaluador AxE	Live	−0.022 (0.407-0.369)	0.647 (0.345-0.827)	0.672 (0.383-0.841)	0.297 (-0.103-0.614)	0.364 (−0.028-0.659)	0.710 (0.444-0.861)	2/6
	S-M	−0.106 (−0.474-0.295)	**0.749 (0.509-0.881)**	**0.771 (0.546-0.892)**	0.470 (0.100-0.726)	0.456 (0.082-0.717)	0.735 (0.485-0.874)	3/6
Evaluador BxC	Live	0.602 (0.279-0.803)	0.722 (0.464-0.867)	**0.883 (0.753-0.947)**	0.393 (0.005-0.678)	0.669 (0.379-0.839)	**0.760 (0.527-0.887)**	5/6
	S-M	0.491 (0.127-0.739)	0.451 (0.076-0.714)	0.723 (0.465-0.868)	0.423 (0.041-0.697)	0.597 (0.272-0.800)	**0.773 (0.550-0.893)**	3/6
Evaluador BxD	Live	0.471 (0.101-0.726)	-----	**0.874 (0.736-0.943)**	0.563 (0.224-0.781)	0.662 (0.369-0.836)	0.481 (0.114-0.732)	3/5
	S-M	0.629	-----	**0.814**	0.447	0.597	0.420	3/5

[12]Adaptada con permiso de MDPI, International Journal of Environmental Research and Public Health, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the Test of Gross Motor Development-Third Edition.* Carballo-Fazanes et al. Copyright (2020).

Evaluadores	Test	Carrera	Desplazamiento lateral†	Galope	Salto con un pie	Salto horizontal	Skip	CCI ≥ 0.5
		(0.318-0.817)		**(0.623-0.914)**	(0.072-0.712)	(0.272-0.800)	(0.038-0.695)	
Evaluador BxE	Live	0.568 (0.231-0.784)	0.529 (0.177-0.761)	**0.816 (0.626-0.915)**	0.350 (-0.044-0.650)	0.492 (0.129-0.739)	0.732 (0.481-0.872)	3/6
	S-M	0.681 (0.399-0.846)	**0.759 (0.526-0.886)**	**0.808 (0.611-0.910)**	0.692 (0.415-0.851)	0.585 (0.255-0.793)	0.608 (0.288-0.806)	6/6
Evaluador CxD	Live	0.569 (0.232-0.784)	-----	**0.827 (0.646-0.920)**	0.656 (0.356-0.832)	**0.784 (0.568-0.899)**	0.636 (0.330-0.822)	5/5
	S-M	0.528 (0.176-0.760)	-----	**0.766 (0.538-0.890)**	0.552 (0.209-0.774)	**0.767 (0.539-0.890)**	0.652 (0.353-0.830)	5/5
Evaluador CxE	Live	0.587 (0.258-0.794)	0.483 (0.117-0.734)	**0.841 (0.672-0.927)**	0.457 (0.084-0.718)	0.407 (0.023-0.687)	0.702 (0.432-0.857)	3/6
	S-M	0.683 (0.401-0.847)	0.640 (0.336-0.824)	0.704 (0.434-0.858)	0.423 (0.042-0.697)	0.591 (0.263-0.796)	0.655 (0.358-0.832)	5/6
Evaluador DxE	Live	0.551 (0.207-0.774)	-----	**0.766 (0.538-0.890)**	0.183 (-0.221-0.533)	0.637 (0.330-0.822)	0.483 (0.117-0.734)	3/5
	S-M	0.459 (0.086-0.719)	-----	**0.826 (0.645-0.920)**	0.433 (0.054-0.703)	0.658 (0.362-0.833)	0.591 (0.263-0.796)	3/5
Todos evaluadores	Live	0.423 (0.243-0.629)	0.494 (0.314-0.686)	**0.795 (0.674-0.889)**	0.431 (0.251-0.636)	0.597 (0.425-0.762)	**0.693 (0.540-0.827)**	3/6
	S-M	0.368 (0.192-0.581)	0.499 (0.319-0.690)	**0.770 (0.641-0.875)**	0.468 (0.287-0.666)	0.616 (0.446-0.775)	0.652 (0.489-0.800)	3/6

*CCI: medidas únicas, consistencia, modelo de efectos aleatorios bidireccional, intervalo de confianza del 95%
Evaluadores A y B: Evaluadores novatos sin formación en Educación Física; Evaluador C: Evaluador novato con formación en Educación Física; evaluadores D y E: evaluadores expertos. † El evaluador D (EE) codificó la habilidad de desplazamiento lateral con puntuaciones máximas en ambas pruebas (en vivo y en cámara lenta); En gris, habilidades con CCI ≥ 0,5; en gris y negrita, habilidades con una fiabilidad inter-observador al menos moderada (CCI y límite inferior ≥ 0,5).

Tabla 4[13]**.** Fiabilidad inter-observador* entre evaluadores. Habilidades de control de objetos

Evaluadores	Test	Bateo	Bote estático	Lanzamiento por encima	Patada	Golpeo con raqueta	Recepción	Lanzamiento por debajo	CCI ≥ 0.5
Evaluador AxB	Live	**0.779 (0.561-0.896)**	0.372 (-0.019-0.664)	0.548 (0.204-0.772)	0.393 (0.006-0.678)	0.632 (0.323-0.819)	0.669 (0.379-0.839)	0.348 (-0.046-0.649)	4/7
	S-M	**0.751 (0.512-0.882)**	**0.778 (0.559-0.896)**	0.695 (0.420-0.853)	0.289 (−0.112-0.609)	0.639 (0.333-0.823)	0.621 (0.306-0.813)	0.598 (0.272-0.800)	6/7
Evaluador AxC	Live	**0.805 (0.607-0.909)**	0.396 (0.009-0.680)	**0.783 (0.567-0.898)**	0.579 (0.246-0.790)	0.569 (0.232-0.784)	**0.876 (0.739-0.943)**	0.589 (0.261-0.796)	6/7
	S-M	**0.788 (0.576-0.901)**	0.631 (0.322-0.819)	0.660 (0.365-0.834)	0.478 (0.110-0.730)	0.674 (0.387-0.842)	**0.929 (0.845-0.968)**	0.475 (0.106-0.729)	5/7
Evaluador AxD	Live	**0.754 (0.517-0.883)**	0.736 (0.488-0.875)	0.586 (0.255-0.793)	0.367 (-0.025-0.661)	0.522 (0.167-0.757)	0.392 (0.005-0.677)	0.594 (0.267-0.798)	5/7
	S-M	0.630 (0.320-0.818)	0.698 (0.426-0.855)	0.445 (0.069-0.710)	0.532 (0.181-0.763)	0.549 (0.204-0.772)	0.389 (0.001-0.675)	0.526 (0.174-0.759)	5/7
Evaluador AxE	Live	0.660 (0.366-0.835)	0.560 (0.219-0.779)	0.570 (0.233-0.785)	0.500 (0.139-0.744)	0.080 (-0.318-0.454)	0.308 (-0.091-0.622)	0.264 (-0.138-0.592)	4/7
	S-M	0.732 (0.480-0.872)	0.545 (0.199-0.770)	0.476 (0.107-0.729)	0.412 (0.028-0.690)	0.378 (-0.012-0.668)	0.297 (-0.103-0.614)	0.322 (-0.076-0.631)	3/7
Evaluador BxC	Live	0.698 (0.426-0.855)	**0.900 (0.787-0.955)**	**0.749 (0.508-0.881)**	0.404 (0.019-0.685)	0.647 (0.346-0.828)	0.600 (0.276-0.801)	0.469 (0.099-0.725)	5/7
	S-M	0.681 (0.398-0.846)	0.735 (0.486-0.874)	**0.776 (0.554-0.895)**	0.462 (0.090-0.721)	0.470 (0.100-0.726)	0.705 (0.436-0.858)	0.614 (0.297-0.810)	5/7
Evaluador BxD	Live	0.544 (0.198-0.770)	0.337 (−0.058-0.642)	0.462 (0.090-0.721)	0.740 (0.493-0.876)	0.691 (0.414-0.851)	0.308 (-0.091-0.622)	0.540 (0.192-0.767)	4/7

[13]Adaptada con permiso de MDPI, International Journal of Environmental Research and Public Health, bajo la licencia Creative Commons Attribution (CC BY) (que permite reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the Test of Gross Motor Development-Third Edition.* Carballo-Fazanes et al. Copyright (2020).

Evaluadores	Test	Bateo	Bote estático	Lanzamiento por encima	Patada	Golpeo con raqueta	Recepción	Lanzamiento por debajo	CCI ≥ 0.5
	S-M	0.582 (0.250-0.791)	0.643 (0.340-0.825)	0.373 (-0.018-0.665)	**0.778 (0.559-0.896)**	0.611 (0.292-0.808)	0.228 (-0.176-0.566)	0.636 (0.330-0.822)	5/7
Evaluador BxE	Live	0.652 (0.354-0.830)	0.385 (−0.004-0.673)	0.596 (0.271-0.800)	0.533 (0.182-0.763)	0.361 (-0.032-0.657)	0.302 (-0.097-0.618)	0.292 (-0.109-0.611)	3/7
	S-M	**0.809 (0.614-0.911)**	0.580 (0.248-0.790)	0.608 (0.288-0.806)	0.534 (0.184-0.764)	0.482 (0.116-0.733)	0.227 (-0.177-0.565)	0.285 (-0.117-0.606)	4/7
Evaluador CxD	Live	0.716 (0.453-0.864)	0.426 (0.045-0.699)	0.610 (0.290-0.807)	0.389 (0.001-0.675)	0.600 (0.276-0.802)	0.410 (0.026-0.689)	0.711 (0.447-0.892)	4/7
	S-M	0.487 (0.121-0.736)	0.593 (0.267-0.798)	0.684 (0.402-0.847)	0.339 (-0.057-0.642)	0.565 (0.226-0.782)	0.313 (-0.086-0.625)	0.665 (0.374-0.837)	4/7
Evaluador CxE	Live	0.682 (0.400-0.846)	0.491 (0.127-0.739)	0.589 (0.261-0.796)	0.412 (0.028-0.690)	0.357 (-0.036-0.655)	0.303 (-0.097-0.618)	0.248 (-0.155-0.580)	2/7
	S-M	0.615 (0.298-0.810)	0.477 (0.109-0.730)	0.570 (0.234-0.785)	0.378 (-0.012-0.668)	0.409 (0.024-0.688)	0.222 (-0.181-0.562)	0.377 (-0.014-0.667)	2/7
Evaluador DxE	Live	0.574 (0.239-0.787)	0.581 (0.249-0.791)	0.505 (0.146-0.747)	0.600 (0.277-0.802)	0.241 (-0.162-0.575)	0.586 (0.255-0.793)	0.176 (-0.228-0.528)	5/7
	S-M	0.530 (0.187-0.762)	0.525 (0.172-0.759)	0.343 (-0.052-0.645)	0.615 (0.297-0.810)	0.287 (-0.114-0.607)	0.586 (0.255-0.793)	0.466 (0.095-0.724)	3/7
Todos evaluadores	Live	**0.686 (0.531-0.823)**	0.518 (0.339-0.705)	0.595 (0.423-0.761)	0.500 (0.320-0.691)	0.470 (0.289-0.667)	0.496 (0.316-0.688)	0.452 (0.271-0.653)	4/7
	S-M	0.652 (0.489-0.800)	0.627 (0.459-0.783)	0.558 (0.381-0.734)	0.498 (0.318-0.689)	0.502 (0.322-0.692)	0.496 (0.316-0.688)	0.506 (0.326-0.695)	5/7

*CCI: medidas únicas, consistencia, modelo de efectos aleatorios bidireccional, intervalo de confianza del 95%.

evaluadores A y B: Evaluadores novatos sin formación en Educación Física; Evaluador C: Evaluador novato con formación en Educación Física; Evaluadores D y E: Evaluadores expertos.

En gris, habilidades con CCI ≥ 0,5; en gris y negrita, habilidades con una fiabilidad inter-observador al menos moderada (ICC y límite inferior ≥ 0,5).



Figura 1[14]**.** Valores del coeficientes de correlación intraclase (CCI) de las subescalas locomotriz y de control de objetos y la puntuación global.

[14]Adaptada con permiso de MDPI, International Journal of Environmental Research and Public Health, bajo la licencia Creative Commons Attribution (CC BY) (que permite

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

La puntuación global de la fiabilidad intra-observador fue, como mínimo, de moderada a excelente. Sin embargo, los EE y el EN-EF alcanzaron niveles de fiabilidad intra-observador más altos que los EN, con excelentes valores de CCI. En este sentido, estudios anteriores también mostraron una fiabilidad intra-observador de buena a excelente o excelente en el TGMD-2 [24-27] y el TGMD-3 [28-32] para las puntuaciones de las subescalas de habilidades locomotrices y de control de objetos. Aunque la fiabilidad intra-observador de los EN fue al menos de moderada a excelente, los EE y el EN-EF alcanzaron como mínimo una fiabilidad de buena a excelente. También se reportaron tasas de acuerdo elevadas en ambas subescalas en el TGMD-2 [24-27] y en el TGMD-3 [28,30-33].

En el análisis riguroso realizado por habilidad, todos los evaluadores alcanzaron una fiabilidad intra-observador de buena a excelente o excelente para el galope, el bateo y la recepción. Sin embargo, se observaron más diferencias entre los evaluadores en los valores del CCI de las habilidades individuales en comparación con la puntuación global o las puntuaciones de las subescalas. En estudios anteriores también se encontraron variaciones en la concordancia de la fiabilidad intra-observador entre los distintos evaluadores [31], especialmente cuando las habilidades se examinan individualmente [10], obteniéndose resultados que van de pobres a excelentes. En el presente estudio, la fiabilidad intra-observador varió de moderada a excelente para los EE y EN-EF en todas las habilidades (excepto el desplazamiento lateral para el evaluador E). La variabilidad de los valores del CCI fue mayor en los EN, con una fiabilidad intra-observador de moderada a buena o de moderada a excelente en la mitad de las habilidades.

En cuanto a la fiabilidad inter-observador, nuestros resultados muestran que el grado de consistencia varía considerablemente entre

reproducir esta publicación siempre que se atribuya a la obra). *Reliability of the Test of Gross Motor Development-Third Edition.* Carballo-Fazanes et al. Copyright (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

los evaluadores en las habilidades del TGMD-3. Como se ha observado, las variaciones en la fiabilidad inter-observador fueron más visibles en el análisis por habilidades. Los resultados del CCI para las puntuaciones de las subescalas locomotriz, control de objetos y puntuaciones globales fueron superiores a 0.5 en casi todas las comparaciones. La fiabilidad inter-observador para las habilidades locomotrices, entre cinco evaluadores, fue de pobre a moderada para la carrera, el desplazamiento lateral y el salto con un pie (tanto en directo como a cámara lenta); y de pobre a buena para el skip a cámara lenta y el salto horizontal (tanto en directo como a cámara lenta); y de moderada a buena para el skip en directo y el galope en ambos modos de visualización. En general, la habilidad con mayor fiabilidad entre los evaluadores fue el galope, que alcanzó un CCI > 0.5 en todas las comparaciones, llegando a una fiabilidad moderada en 15/20 comparaciones entre evaluadores. Por el contrario, los resultados más bajos de fiabilidad se encontraron para la habilidad de salto con un pie y carrera. Kim et al. [34] también informaron de que el galope es la habilidad con resultados de fiabilidad más altos, y la carrera la que tiene los coeficientes de fiabilidad más bajos. Por el contrario, Maeng et al. [31] y Ayán et al. [35] obtuvieron la fiabilidad más robusta en el salto con un pie y la más débil en el galope, al contrario de nuestros resultados.

En lo que respecta a las habilidades de control de objetos, la fiabilidad inter-observador fue de baja a moderada para el bote estático (en directo), el lanzamiento por encima de la cabeza (a cámara lenta), la patada, el golpeo con raqueta, la recepción y el lanzamiento por debajo (tanto en directo como a cámara lenta); de baja a buena para el bateo (a cámara lenta), el bote estático (a cámara lenta) y el lanzamiento por encima (en directo); y de moderada a buena para el bateo. En general, el bateo fue la habilidad con mejor fiabilidad entre los evaluadores, alcanzando un CCI > 0.5 y una fiabilidad moderada en 19/20 comparaciones. Los coeficientes de fiabilidad más bajos se alcanzaron en la habilidad de recepción. Este resultado está en

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

concordancia con estudios anteriores [36] pero, de nuevo, en contra de otros que encontraron la recepción como la habilidad con mayor fiabilidad entre los evaluadores [10].

Nuestros resultados muestran una menor concordancia en las habilidades locomotrices en comparación con las habilidades de control de objetos. Rintala et al. [10] y Palmer y Brian [22], que informaron de resultados similares en el TGMD-2, sugirieron que antes de utilizar el TGMD, puede que sea necesario un mayor entrenamiento con las habilidades locomotrices, nuestros resultados apoyan esta afirmación. Por otra parte, parece que los antecedentes y la experiencia de los evaluadores no influyeron en la fiabilidad inter-observador, teniendo en cuenta que la fiabilidad varió no sólo entre EN y EE, sino también entre EN y EE entre sí mismos. Por el contrario, Palmer y Brian [22] encontraron que los evaluadores novatos no podían alcanzar un acuerdo significativo con los evaluadores expertos que evaluaban la prueba TGMD-2. No obstante, se ha mostrado que los evaluadores novatos alcanzaron un acuerdo significativamente similar al de los expertos al evaluar la calidad de movimiento con otros test como el Landing Error Scoring System [37] y Functional Movement Screen [38].

La variación encontrada en la fiabilidad inter-observador en nuestro estudio podría deberse a las diferentes interpretaciones de cada uno de los criterios de rendimiento de las habilidades. Algunos de ellos dan lugar a la subjetividad y varían según la interpretación concreta de los evaluadores. Por ejemplo, en la habilidad de salto con un pie, el segundo criterio, "*El pie de la pierna libre permanece detrás del cuerpo*" podría interpretarse de forma diferente; podría considerarse que se puntúa con 1 si el pie libre siempre permanece detrás del cuerpo. Sin embargo, también podría considerarse que puntúa 1 si el niño realizase el criterio correctamente la mayor parte del tiempo. Por lo tanto, se hace necesario obtener acuerdos previos a la evaluación entre los evaluadores o profesionales que pretenden evaluar las HMB en niños. Se ha sugerido que puede ser necesario un mayor

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

entrenamiento para aquellas habilidades más difíciles de evaluar e incluso que se podrían redefinir algunos de los criterios para hacerlos más objetivos [10,20,31,34,36]; nuestros resultados aportan más apoyo a esta recomendación.

Además, una reciente revisión sistemática sobre la fiabilidad del TGMD-2 y 3 mostró valores de CCI al menos entre buenos y excelentes para la fiabilidad inter-observador en ambas subescalas y en la puntuación global, superiores a los obtenidos en este estudio. Sin embargo, estas diferencias podrían estar relacionadas con la metodología adoptada o el análisis estadístico utilizado [20]. En el caso del CCI, utilizado en muchos estudios, existen varias razones que dificultan la comparación de los resultados. En primer lugar, los investigadores tienen que elegir un modelo de CCI (modelo de efectos aleatorios unidireccional, modelo de efectos aleatorios bidireccional o modelo de efectos mixtos bidireccional), la definición de CCI (acuerdo absoluto o consistencia) y el tipo de CCI (medidas únicas o medias promedio) según el estudio concreto. En segundo lugar, no hay valores estándar para informar de la fiabilidad. En tercer lugar, algunos estudios no informaron del modelo/tipo/definición del CCI o informaron de la fiabilidad basándose en los valores del CCI sin mencionar el intervalo de confianza. Nuestro estudio siguió el diagrama de flujo propuesto por Koo et al. [23] para seleccionar el modelo, el tipo y la definición. Muchos estudios han utilizado diferentes modelos, en la mayoría de los casos, menos exigentes, lo que podría explicar los resultados de fiabilidad más altos obtenidos en comparación con nuestro estudio.

En cuanto a los valores del CCI y los intervalos de confianza del 95%, la interpretación también se basó en las recomendaciones de Koo et al. [23]. Respecto a esto, los valores del CCI no describen la fiabilidad en sí mismos, sino que es necesario expresar la fiabilidad en términos de los límites inferior y superior del intervalo de confianza. Por último, consideramos que la fiabilidad es excelente si se obtiene un CCI igual o superior a 0.9, mientras que otras valoraciones lo consideran

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

superior a 0.75 [39,40]. Todo esto dificulta la comparación de la fiabilidad utilizada en diferentes estudios, ya que esta variabilidad provoca un problema de interpretación.

Limitaciones

Este estudio no está exento de limitaciones. Nuestro estudio de fiabilidad se realizó evaluando a 25 niños sanos. La muestra era pequeña, por lo que los resultados deben interpretarse con precaución. Además, intentamos simular la evaluación en directo mostrando los vídeos sólo una vez y reproduciéndolos a velocidad normal. Este procedimiento no representa con precisión la realidad de una evaluación en directo por parte de los profesionales que tienen que puntuar el rendimiento de los niños. Sin embargo, desde el punto de vista de la investigación, es una forma útil para que diferentes evaluadores puedan realizar la evaluación en cualquier momento.

5. Conclusiones

El TGMD-3 mostró una fiabilidad intra-observador de moderada a excelente para la puntuación global y las subescalas de habilidades locomotrices y de control de objetos, y una fiabilidad inter-observador de moderada a buena para las puntuaciones globales y de la subescala de habilidades de control de objetos, y de pobre a buena para la subescala locomotriz. Los evaluadores expertos y el evaluador novato con formación en educación física lograron una mayor fiabilidad intra-observador que los evaluadores novatos, no obstante, la fiabilidad inter-observador no pareció estar influenciada por la formación y la experiencia de los evaluadores. Los modos de visualización tampoco parecieron influir en la fiabilidad, pero es necesario seguir investigando en esta línea, teniendo en cuenta las limitaciones del estudio. Se encontró una mayor variabilidad en la fiabilidad intra- e inter-observador cuando se analizaron las habilidades por separado.

En conclusión, nuestros mayores resultados de fiabilidad intra-observador en comparación con la inter-observador, sugieren que los

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

criterios de cada habilidad pueden ser interpretados de forma diferente por los evaluadores. Para la práctica clínica, sería recomendable que los evaluadores llegaran a un acuerdo antes de la evaluación para evitar interpretaciones subjetivas que pudieran distorsionar los resultados. Además, podría ser necesaria una revisión de algunos criterios para permitir la replicación de la investigación.

REFERENCIAS

1. Gallahue, D.L.; Donnelly, F.C. *Developmental Physical Education for All Children*, 4th ed.; Human Kinetics: Champaing, IL, USA, 2007.
2. Gallahue, D.; Ozmun, J.; Goodway, J. *Understanding Motor Development: Infants, Children, Adolescents, Adults*, 7th ed.; McGraw-Hill: New York, NY, USA, 2012.
3. Logan, S.W.; Ross, S.M.; Chee, K.; Stodden, D.F.; Robinson, L.E. Fundamental motor skills: A systematic review of terminology. *J. Sports Sci.* **2018**, *36*, 781–796, doi:10.1080/02640414.2017.1340660.
4. Lubans, D.R.; Morgan, P.J.; Cliff, D.P.; Barnett, L.M.; Okely, A.D. Fundamental Movement Skills in Children and Adolescents. Review of associated health benefits. *Sports Med.* **2010**, *40*, 1019–1035, doi:10.2165/11536850-000000000-00000.
5. Cattuzzo, M.T.; Henrique, R.D.S.; Ré, A.H.N.; De Oliveira, I.S.; Melo, B.M.; Moura, M.D.S.; De Araújo, R.C.; Stodden, D. Motor competence and health related physical fitness in youth: A systematic review. *J. Sci. Med. Sport* **2016**, *19*, 123–129, doi:10.1016/j.jsams.2014.12.004.
6. Barnett, L.M.; Lai, S.K.; Veldman, S.L.C.; Hardy, L.L.; Cliff, D.P.; Morgan, P.J.; Zask, A.; Lubans, D.R.; Shultz, S.P.; Ridgers, N.D.; et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Med.* **2016**, *46*, 1663–1688, doi:10.1007/s40279-016-0495-z.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7. Robinson, L.E.; Stodden, D.F.; Barnett, L.M.; Lopes, V.P.; Logan, S.W.; Rodrigues, L.P.; D'Hondt, E. Motor Competence and its Effect on Positive Developmental Trajectories of Health. *Sports Med.* **2015**, *45*, 1273–1284, doi:10.1007/s40279-015-0351-6.
8. Haapala, E.A. Cardiorespiratory Fitness and Motor Skills in Relation to Cognition and Academic Performance in Children—A Review. *J. Hum. Kinet.* **2013**, *36*, 55–68, doi:10.2478/hukin-2013-0006.
9. Costa, H.J.T.; Barcala-Furelos, R.; Abelairas-Gómez, C.; Arufe-Giráldez, V. The Influence of a Structured Physical Education Plan on Preschool Children's Psychomotor Development Profiles. *Australas. J. Early Child.* **2015**, *40*, 68–77, doi:10.1177/183693911504000209.
10. Rintala, P.; Sääkslahti, A.; Iivonen, S. Reliability Assessment of Scores from Video-Recorded TGMD-3 Performances. *J. Mot. Learn. Dev.* **2017**, *5*, 59–68, doi:10.1123/jmld.2016-0007.
11. Logan, S.W.; Robinson, L.E.; Rudisill, M.E.; Wadsworth, D.D.; Morera, M. The comparison of school-age children's performance on two motor assessments: The Test of Gross Motor Development and the Movement Assessment Battery for Children. *Phys. Educ. Sport Pedagog.* **2014**, *19*, 48–59, doi:10.1080/17408989.2012.726979.
12. Piek, J.P.; Dawson, L.; Smith, L.M.; Gasson, N. The role of early fine and gross motor development on later motor and cognitive ability. *Hum. Mov. Sci.* **2008**, *27*, 668–681, doi:10.1016/j.humov.2007.11.002.
13. Skinner, R.A.; Piek, J.P. Psychosocial implications of poor motor coordination in children and adolescents. *Hum. Mov. Sci.* **2001**, *20*, 73–94, doi:10.1016/s0167-9457(01)00029-x.
14. Burton, A.; Miller, D. *Movement Skill Assessment*; Human Kinetics: Champaing, IL, USA, 1998.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

15. Yun, J.; Shapiro, D.R. A Quantitative Approach to Movement Skill Assessment for Children with Mental Retardation. *Adapt. Phys. Act. Q.* **2004**, *21*, 269–280, doi:10.1123/apaq.21.3.269.
16. Clark, J.E. Motor Development. In *Encyclopedia of Human Behavior*, 3rd ed.; Ramachandran, V.S., Ed.; Academic Press: New York, NY, USA, 1994; pp. 245–255.
17. Payne, V.G.; Isaacs, L.D. *Human Motor Development*; McGraw-Hill: New York, NY, USA, 2008.
18. Ulrich, D.A. *Test of Gross Motor Development*, 2nd ed.; PRO-ED: Austin, TX, USA, 2000.
19. Ulrich, D.A. *Test of Gross Motor Development*, 3rd ed.; PRO-ED: Austin, TX, USA, 2019.
20. Rey, E.; Carballo-Fazanes, A.; Varela-Casal, C.; Abelairas-Gómez, C. Reliability of the test of gross motor development: A systematic review. *PLoS ONE* **2020**, *15*, e0236070, doi:10.1371/journal.pone.0236070.
21. Akuffo, P.B.; Hodge, S.R. Roles and Responsibilities of Adapted Physical Education Teachers in an Urban School District. *Educ. Urban Soc.* **2008**, *40*, 243–268, doi:10.1177/0013124507304132.
22. Palmer, K.K.; Brian, A. Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. *J. Mot. Learn. Dev.* **2016**, *4*, 142–151, doi:10.1123/jmld.2015-0035.
23. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163, doi:10.1016/j.jcm.2016.02.012.
24. Aye, T.; Oo, K.S.; Khin, M.T.; Kuramoto-Ahuja, T.; Maruyama, H. Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. *J. Phys. Ther. Sci.* **2017**, *29*, 1726–1731, doi:10.1589/jpts.29.1726.
25. Capio, C.M.; Eguia, K.F.; Simons, J. Test of gross motor development-2 for Filipino children with intellectual disability:

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Validity and reliability. *J. Sports Sci.* **2015**, *34*, 10–17, doi:10.1080/02640414.2015.1033643.

26. Farrokhi, A.; Zareh, Z.; Karimi, A.; Kazemnejad, A.; Ilbeigi, S. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3–10 aged children of Tehran City. *J. Phys. Educ. Sport Manag.* **2014**, *5*, 18–28, doi:10.5897/JPESM12.003.
27. Houwen, S.; Hartman, E.; Jonker, L.; Visscher, C. Reliability and Validity of the TGMD-2 in Primary-School-Age Children with Visual Impairments. *Adapt. Phys. Act. Q.* **2010**, *27*, 143–159, doi:10.1123/apaq.27.2.143.
28. Allen, K.A.; Bredero, B.; Van Damme, T.; Ulrich, D.A.; Simons, J. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. *J. Autism Dev. Disord.* **2017**, *47*, 813–833, doi:10.1007/s10803-016-3005-0.
29. Estevan, I.; Molina-García, J.; Queralt, A.; Álvarez, O.; Castillo, I.; Barnett, L. Validity and Reliability of the Spanish Version of the Test of Gross Motor Development–3. *J. Mot. Learn. Dev.* **2017**, *5*, 69–81, doi:10.1123/jmld.2016-0045.
30. Mohammadi, F.; Bahram, A.; Khalaji, H.; Ulrich, D.; Ghadiri, F. Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. *J. Mot. Learn. Dev.* **2019**, *7*, 106–121, doi:10.1123/jmld.2017-0045.
31. Maeng, H.; Webster, E.K.; Pitchford, E.; Ulrich, D. Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition among Experienced TGMD-2 Raters. *Adapt. Phys. Act. Q.* **2017**, *34*, 442–455, doi:10.1123/apaq.2016-0026.
32. Valentini, N.C.; Zanella, L.W.; Webster, E.K. Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. *J. Mot. Learn. Dev.* **2017**, *5*, 15–28, doi:10.1123/jmld.2016-0002.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

33. Wagner, M.O.; Webster, E.K.; Ulrich, D.A. Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation): Results of a Pilot Study. *J. Mot. Learn. Dev.* **2017**, *5*, 29–44, doi:10.1123/jmld.2016-0006.
34. Kim, Y.; Park, I.; Kang, M. Examining Rater Effects of the TGMD-2 on Children with Intellectual Disability. *Adapt. Phys. Act. Q.* **2012**, *29*, 346–365, doi:10.1123/apaq.29.4.346.
35. Ayán, C.; Cancela, J.; Sanchez-Lastra, M.A.; Carballo-Roales, A.; Domínguez-Meis, F.; Redondo-Gutiérrez, L. Fiabilidad y Validez de la Batería TGMD-2 en Población Española. *Rev. Iberoam. Diagnóstico Eval.* **2019**, *50*, 21–34, doi:10.21865/ridep50.1.02.
36. Barnett, L.M.; Minto, C.; Lander, N.; Hardy, L.L. Interrater reliability assessment using the Test of Gross Motor Development-2. *J. Sci. Med. Sport* **2014**, *17*, 667–670, doi:10.1016/j.jsams.2013.09.013.
37. Onate, J.; Cortes, N.; Welch, C.; Van Lunen, B.L. Expert Versus Novice Interrater Reliability and Criterion Validity of the Landing Error Scoring System. *J. Sport Rehabil.* **2010**, *19*, 41–56, doi:10.1123/jsr.19.1.41.
38. Minick, K.I.; Kiesel, K.B.; Burton, L.; Taylor, A.; Plisky, P.; Butler, R.J. Interrater Reliability of the Functional Movement Screen. *J. Strength Cond. Res.* **2010**, *24*, 479–486, doi:10.1519/jsc.0b013e3181c09c04.
39. Fleiss, J.L. *Design and Analysis of Clinical Experiments*; John Wiley & Sons: New York, NY, USA, 1986.
40. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290, doi:10.1037/1040-3590.6.4.284.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.1.3 Artículo 3: Fiabilidad inter-observador del Test of Gross Motor Development – Third Edition, tras el consenso previo de los evaluadores en los criterios de rendimiento

[Artículo en revisión]

Aida Carballo-Fazanes[1,2] (primera autora)

Ezequiel Rey[3] (segundo autor)

Nadia C. Valentini[4] (tercera autora)

Cristina Varela-Casal[2,3] (cuarta autora)

Cristian Abelairas-Gómez[1,2,5] (último autor)

[1] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, España.

[2] Grupo de Investigación CLINURSID, Departamento de Psiquiatría, Radiología, Salud Pública, Enfermería y Medicina, Universidade de Santiago de Compostela, España.

[3] Grupo de investigación REMOSS, Facultad de Educación y Ciencias del Deporte, Universidade de Vigo, España.

[4] Departmento de Educación Física, Fisioterapia y Danza, Universidade Federal do Rio Grande do Sul, Brazil.

[5] Facultad de Ciencias de la Educación, Universidade de Santiago de Compostela, España.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen

En este estudio se calculó la fiabilidad inter-observador del Test of Gross Motor Development - Tercera Edición (TGMD-3) después de que los evaluadores llegaran a un consenso sobre los criterios de rendimiento de cada habilidad. Tres evaluadores evaluaron las habilidades motrices básicas (HMB) de los mismos 25 niños con el TGMD-3 en dos momentos diferentes, utilizando dos métodos diferentes: (i) una vez siguiendo exactamente los criterios de rendimiento recogidos en el manual del TGMD-3; y (ii) de nuevo, tras un período de lavado de 9 meses, después de que los evaluadores llegaran a un consenso para los criterios de rendimiento de cada habilidad. Después de calcular y comparar la fiabilidad entre los tres evaluadores en estos dos momentos y métodos de evaluación, se encontró una mejora en la fiabilidad inter-observador después del consenso de los evaluadores en las habilidades locomotrices (fiabilidad de moderada a buena en 2 de las 6 habilidades inicialmente y al menos de moderada a excelente en 4 de las 6 habilidades después del consenso) y en las de control de objetos (fiabilidad de moderada a buena en 1 de las 7 habilidades inicialmente y al menos de moderada a excelente en 4 de las 7 habilidades después de la construcción de criterios de consenso). En las puntuaciones de las subescalas y en la puntuación global, los evaluadores alcanzaron una fiabilidad al menos entre moderada y buena en su segunda evaluación posterior al consenso. Basándonos en esta mejora de la fiabilidad tras el consenso, recomendamos que los investigadores incluyan el consenso de los evaluadores sobre los criterios de rendimiento antes de evaluar las HMB de los niños y niñas o de orientar las intervenciones curriculares en Educación Física a partir de los datos del TGMD-3.

Palabras clave: Evaluación; desarrollo infantil; competencia motriz; habilidades motrices gruesas; TGMD.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. Introducción

La evaluación de las habilidades motrices básicas (HMB) de los niños y niñas ha ganado importancia en las últimas décadas (1,2). La evaluación de las HMB puede ser valiosa para identificar a los niños con bajos niveles de competencia motriz, permitir comparaciones de los niveles de competencia motriz entre diferentes poblaciones, o guiar las intervenciones escolares adecuadas para promover el desarrollo saludable de los niños (2,3).

Existen una serie de herramientas de evaluación de las HMB con fines clínicos, educativos y de investigación (4). Estas herramientas pueden ser ampliamente clasificadas en (a) pruebas orientadas a la cantidad/producto, aquellas que ofrecen mediciones cuantificables del producto o resultado de los movimientos de los niños (p. ej.: la distancia saltada); (b) pruebas orientadas a la calidad/proceso, aquellas que miden la calidad del proceso de movimiento para determinar si los niños han alcanzado algunos criterios de rendimiento predefinidos (p. ej.: valoración sobre la calidad del movimiento del brazo durante el salto); y (c) pruebas híbridas con métodos de puntuación que combinan ambos enfoques (1). Las pruebas orientadas al proceso pueden proporcionar información cualitativa valiosa para guiar la enseñanza de los niños en la realización de un nuevo movimiento motriz (5). Sin embargo, la puntuación de un observador sobre los movimientos de los niños puede ser compleja, y puede requerir que los evaluadores tengan un amplio conocimiento de las HMB o un entrenamiento de habilidades específicas en su evaluación precisa (6).

Entre las herramientas de evaluación orientadas al proceso, el Test of Gross Motor Development-Tercera Edición (TGMD-3 (7) (TGMD), y sus predecesores TGMD y TGMD-2, se basan en las observaciones de los evaluadores (1) que se organizan en dos subescalas - habilidades locomotrices y de control de objetos. Una revisión sistemática reciente sugirió que el TGMD, en sus diferentes ediciones, es la herramienta utilizada más frecuentemente para medir la competencia en HMB de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

los niños en entornos educativos, de investigación y clínicos (8). Un prerrequisito para el uso de cualquier herramienta de evaluación es que muestre una adecuada fiabilidad y validez (9,10), y, a pesar de destacar estas fortalezas psicométricas del TGMD-3, Rey et al. (8) también encontraron en su revisión que los valores de fiabilidad inter-observador eran generalmente más bajos que los observados para la fiabilidad intra-observador, probablemente debido a la dificultad de los evaluadores para lograr un consenso en la interpretación de los criterios de puntuación para algunos de los componentes de las habilidades (11,12). Las diferencias entre los evaluadores están relacionadas con los distintos puntos de vista, interpretaciones y métodos de evaluación. El manual del TGMD-3 proporciona instrucciones para calificar si un niño cumple con ciertos criterios de rendimiento; pero estos juicios siempre estarán sujetos a la discreción de cada evaluador (13), a menos que haya habido un consenso entre evaluadores que, idealmente, sea ampliamente difundido y compartido.

Barnett et al. (11) examinaron la fiabilidad inter-observador de la subescala de control de objetos TGMD-2 mediante observación en directo y encontraron algunas definiciones problemáticas y difíciles de identificar de los criterios de rendimiento que deben ser aclaradas y discutidas entre los evaluadores en un proceso de consenso. Esta debilidad del TGMD podría implicar la necesidad de una medición más precisa obtenida mediante la reducción del sesgo de subjetividad de cada evaluador con el consenso de criterios antes de la evaluación. Incluso después de lograr un acuerdo satisfactorio entre evaluadores, surgen preguntas sobre cómo se mantiene a lo largo del tiempo y cuándo puede ser necesario renovar la formación y la creación de este consenso. Esta información es relevante, especialmente en los estudios en los que las mediciones de las HMB de los niños abarcan varios períodos, como en la intervención y la investigación longitudinal. Sin embargo, según el conocimiento de los autores y autoras, ningún estudio ha comparado la fiabilidad inter-observador del TGMD

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

después de que los evaluadores hayan llegado a un consenso entre ellos sobre los criterios de rendimiento de las HMB. Por lo tanto, nuestro objetivo en este estudio fue determinar la fiabilidad inter-observador del TGMD-3 después de que los evaluadores llegasen a su propio consenso sobre los criterios de rendimiento.

2. Métodos

2.1 Participantes

Veinticinco niños y niñas sanos de primaria participaron en este estudio (15 niñas, 10 niños; *M edad*= 9.16, *DT* = 1.31 años). Se obtuvo el consentimiento informado por escrito de todos los padres/madres o tutores legales de los niños, así como el consentimiento verbal de los participantes. Este estudio siguió los principios éticos de la Declaración de Helsinki y fue aprobado por el Comité Ético de la Facultad de Ciencias de la Educación y el Deporte (Universidade de Vigo).

2.2 Diseño del estudio

En una primera fase, cinco evaluadores utilizaron 13 habilidades del TGMD-3 grabadas en vídeo y realizadas por 25 participantes para evaluar su competencia motriz (14). Inicialmente, se explicó cada habilidad a los participantes. A continuación, los participantes vieron un vídeo, a velocidad normal y a cámara lenta, realizado por el autor del TGMD, en el que se mostraba la ejecución correcta de las habilidades (15). A continuación, los participantes realizaron tres ensayos de cada habilidad: el primero fue un ensayo de práctica (no puntuado) para demostrar la comprensión de la habilidad, y los otros dos fueron grabados en vídeo (cámara Nikon D5300) para que las HMB pudieran ser puntuadas por los evaluadores posteriormente. Se estudió la fiabilidad intra-observador e inter-observador comparando las puntuaciones de los evaluadores novatos con las de los expertos al ver el vídeo de forma continua sin pausas a velocidad normal y visualizando el vídeo a cámara lenta tantas veces como el evaluador

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

necesitara (14). La fiabilidad intra-observador fue mayor que la fiabilidad inter-observador (ambas calculadas a través del coeficiente de correlación intra-clase (CCI)), independientemente de la experiencia del evaluador o del modo de visionado, presumiblemente debido a los distintos puntos de vista de los evaluadores individuales (14).

En una segunda fase de esta investigación, llevada a cabo nueve meses después, tres de los cinco evaluadores iniciales llegaron a un acuerdo sobre los criterios de rendimiento de cada habilidad en el TGMD-3 tras un proceso de elaboración de consenso entre evaluadores. Después, los tres evaluadores evaluaron las mismas 13 habilidades grabadas en vídeo de los mismos 25 participantes incluidos en la primera fase del estudio.

2.3 Instrumento de evaluación - Test of Gross Motor Development-Third Edition [TGMD-3]

El TGMD-3 es una batería de pruebas orientada al proceso para evaluar el rendimiento de las habilidades motrices gruesas de los niños (de 3 a 10 años y 11 meses) (7). Está organizada en dos subescalas, que miden las habilidades locomotrices y las habilidades de control de objetos. La subescala locomotriz mide las habilidades que requieren movimientos coordinados direccionales (carrera, galope, salto con un pie, skip, salto horizontal y desplazamiento lateral). La subescala de habilidades de control de objetos mide las habilidades relacionadas con la interceptación y la propulsión de objetos (bateo, bote estático, lanzamiento por encima, patada, golpeo con raqueta, recepción y lanzamiento por debajo). Cada destreza incluye de 3 a 6 criterios de rendimiento que se puntúa como "0" o "1", dependiendo de la ausencia o presencia del criterio. Así, se obtiene una puntuación para cada habilidad, para cada subescala, y la suma de estas habilidades de los ítems de ambas subescalas da lugar a la puntuación global del test.

Como se ha señalado, tras una descripción verbal y una demostración práctica basada en un vídeo, los participantes realizaron tres ensayos

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

de cada habilidad, siendo el primero un ensayo de práctica no puntuado para demostrar la comprensión del niño de la habilidad, y los otros dos ensayos puntuados por el evaluador, en los que "1" indicaba una ejecución correcta y "0" una ejecución incorrecta. Las puntuaciones iban de 0 a 46 puntos para la subescala de locomotriz y de 0 a 54 para la subescala de habilidad de control de objetos, con una puntuación máxima global de 100.

2.4 Creación del consenso sobre los criterios del TGMD-3 entre los evaluadores

Entre la primera y segunda evaluación (después de un período de lavado de 9 meses) de los mismos participantes utilizando el TGMD-3, los tres evaluadores debatieron sobre los criterios de rendimiento de cada habilidad llegando a un consenso sobre la mejor manera de calificarlos. Sus evaluaciones iniciales (primera fase) se basaron en las directrices del manual del TGMD-3 para calificar las 13 habilidades. En la segunda fase, los evaluadores hicieron revisiones individuales de los criterios de rendimiento para las 13 habilidades y luego consideraron todos sus comentarios por separado, habilidad por habilidad, y, a través del debate conjunto, establecieron acuerdos generales y específicos (Tabla 1 y Tabla 2). Llegaron a acuerdos generales sobre factores que podían aplicarse globalmente a todas las habilidades, y a acuerdos específicos que sólo eran aplicables a una habilidad concreta.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 1. Consenso alcanzado respecto a las habilidades locomotrices. [Elaboración propia]

Consenso general para todas las habilidades

1. En las habilidades rítmicas con varios patrones (p. ej.: el galope), el criterio de ejecución se puntuará con "1" cuando se mantenga durante la mitad (patrones pares) o la mitad +1 (patrones impares) de los patrones.
2. Si el evaluador tiene dudas en la valoración de un criterio de ejecución se puntuará con "1".

Consenso específico para las habilidades de la subescala locomotriz

Habilidad	Criterio de rendimiento	Consenso
Galope	1. Brazos flexionados y levantados a la altura de la cintura en el despegue	Tener en cuenta cada despegue, no solo el primero.
	2. Un paso adelante con el pie delantero seguido de un paso con el pie trasero hasta una posición adyacente o detrás del pie adelantado (no delante)	Si el pie trasero supera claramente el pie delantero, este criterio se puntuará como "0".
	3. Breve período en el que ambos pies están en el aire	----
	4. Mantiene un patrón rítmico durante 4 galopes consecutivos	Si los criterios 1, 2 y 3 tienen una puntuación de "0", el criterio 4 también tendrá una puntuación de "0".
Desplazamiento lateral	1. El cuerpo girado de lado de manera que los hombros estén alineados con la línea del suelo (se puntúa solo el lado dominante)	El cuerpo tiene que estar alineado con la dirección de movimiento.
	2. Un paso de lado con el pie delantero seguido de un deslizamiento con el pie trasero hasta un punto próximo al pie delantero (se puntúa solo el lado dominante)	Aunque el pie trasero llegue a tocar el delantero, este criterio se puntúa como "1".
	3. Un mínimo de 4 ciclos paso-deslizamiento hacia la derecha	----
	4. Un mínimo de 4 ciclos paso-deslizamiento hacia la izquierda	----
Salto con un pie	1. La pierna libre se balancea delante de forma pendular para hacer fuerza	La rodilla de la pierna que no apoya (pierna libre) debe sobrepasar la línea del trinco con un movimiento pendular.
	2. El pie de la pierna libre permanece detrás del cuerpo (no se cruza por delante)	Si el pie de la pierna libre supera claramente al de la pierna que apoya, este criterio se puntuará como "0".
	3. Brazos flexionados y balanceándose hacia delante para hacer fuerza	----
	4. Despega y aterriza 4 veces consecutivas con el pie dominante	----

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Skip	1. Un paso hacia delante seguido de un salto sobre el mismo pie	----
	2. Los brazos se flexionan y se mueven en oposición a las piernas para producir fuerza	----
	3. Completa 4 saltos rítmicos y consecutivos alternando las piernas	----
Salto horizontal	1. El movimiento preparatorio incluye la flexión de ambas rodillas y los brazos extendidos detrás del cuerpo	Los brazos extendidos detrás cuerpo tomando como referencia el glúteo.
	2. Brazos que se extienden con fuerza delante y arriba alcanzando la extensión completa por encima de la cabeza	No es necesario alcanzar una extensión completa de brazos. Si balancea los brazos hacia delante y la mano pasa por encima de la cabeza, este criterio se puntuará como "1".
	3. Despegue y aterrizaje con ambos pies simultáneos	----
	4. Brazos empujando hacia abajo durante el aterrizaje	Durante el aterrizaje, los brazos deben balancearse hacia abajo, no es necesario que el balanceo sea por detrás del tronco. Si el criterio 2 se puntúa como "0", este criterio también.
Carrera	1. Los brazos se mueven en oposición a las piernas, codos flexionados	Si en alguno de los criterios se falla más de tres veces, ese criterio se puntuará como "0". ** En este caso no se aplica la norma de la mitad de patrones correctos.*
	2. Breve período donde ambos pues están en el aire	
	3. El contacto del pie se reduce al talón o punta (no con pie plano)	
	4. Pierna libre flexionada a 90º cerca del glúteo	

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 2. Consenso alcanzado respecto a las habilidades de control de objetos. [Elaboración propia]

Consenso general para todas las habilidades

1. En las habilidades rítmicas con varios patrones (p. ej.: el galope), el criterio de ejecución se puntuará con "1" cuando se mantenga durante la mitad (patrones pares) o la mitad +1 (patrones impares) de los patrones.
2. Si el evaluador tiene dudas en la valoración de un criterio de ejecución se puntuará con "1".

Consenso específico para las habilidades de la subescala de control de objetos

Habilidad	Criterio de rendimiento	Consenso
Bateo	1. La mano dominante agarra el bate por encima de la no dominante	
	2. El lado no dominante mira hacia el lado del lanzador imaginario con los pies paralelos	Alineación: pies-cadera-hombro y orientando hacia zona de lanzamiento.
	3. La cadera y el hombro rotan y se suspenden durante el balanceo	Fijarse especialmente en la cadera. Si el criterio 2, es puntuado como "0", este criterio también.
	4. Transfiere el peso corporal hacia el pie adelantado	Si el criterio 3, es puntuado como "0", este criterio también.
	5. El bate golpea la pelota enviándola hacia adelante	
Patada	1. Aproximación rápida y continuada al balón	No caminar y realizar una aproximación continua.
	2. Paso alargado o zancada antes del golpeo del balón	-----
	3. Pie de apoyo colocado cerca del balón	-----
	4. Golpea el balón con el empeine o el interior del pie dominante (no con los dedos)	Este criterio se puntuará como "0" si claramente el balón es golpeado de puntillas o con el lateral.
Golpeo con raqueta	1. Realiza un movimiento de retroceso con la raqueta cuando la pelota bota	Si el hombro comienza en abducción (raqueta alejada), este se criterio se puntuará como "o" porque no hay retroceso.
	2. Camina hacia la pelota con el pie no dominante	-----
	3. Golpea la pelota hacia la pared	La dirección de la pelota deberá ser hacia donde tiene orientado el hombro.
	4. La raqueta se dirige hacia el hombro no dominante	El movimiento de la raqueta no se detiene justo después del golpeo. Continúa hacia el hombro no dominante.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Lanzamiento por encima	1. El armado del brazo se inicia con un movimiento hacia abajo del brazo/mano	Movimiento hacia abajo, la mano debe pasar por detrás de la cabeza.
	2. Rotación de caderas y hombro hacia el lado que no lanza	Si realiza el lanzamiento con ambos pies en paralelo, este criterio se puntuará como "0".
	3. El peso se transfiere dando un paso con el pie contrario al brazo de lanzamiento	Si empieza con el pie contrario al brazo de lanzamiento adelantado, este criterio se puntuará como "0".
	4. El movimiento continúa después de soltar la bola diagonalmente hacia la cadera del brazo que no lanza	-----
Lanzamiento por debajo	1. La mano dominante balancea hacia atrás-abajo llegando detrás del tronco	-----
	2. Dar un paso adelante con el pie contrario a la mano que lanza	Si empieza con el pie contrario al brazo de lanzamiento adelantado, este criterio se puntuará como "0".
	3. Lanza el balón hacia delante golpeando la pared sin rebote	-----
	4. La mano llega al menos al nivel del pecho después del lanzamiento	-----
Recepción	1. Fase de preparación con los brazos delante del cuerpo y codos flexionados	No es necesario que os codos estén flexionados a 90º. Los brazos deben sobrepasar el tronco con una ligera flexión de los codos.
	2. Los brazos se extienden mientras alcanzan la pelota al llegar	Los brazos se mueven hacia delante mirando el balón.
	3. La pelota se coge solo con las manos	-----
Bote estático	1. Contacta el balón con la mano a la altura de la cintura	Si el contacto se realiza por encima o debajo de la cintura, este criterio se puntuará como "0".
	2. Empuja el balón con la yema de los dedos sin dar golpes (no usa las palmas)	El balón debe ser empujado, no golpeado. Si el balón es golpeado, este criterio se puntuará como "0".
	3. Mantiene el control de la pelota durante 4 botes sin mover los pies para conseguirlo	Si los pies ese mueven durante la recepción, este criterio se puntuará como "1" porque es posterior al cuarto bote.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2.5 Análisis estadístico

Se utilizó el CCI para evaluar la fiabilidad inter-observador. Siguiendo a Koo y Li (16), se basaron los valores del CCI y sus intervalos de confianza del 95% en un modelo de medida única (tipo), consistencia (definición) y efectos aleatorios bidireccionales (modelo). Los valores inferiores a 0.50 indicaron una fiabilidad pobre, los valores entre 0.50 y 0.75 indicaron una fiabilidad moderada, los valores entre 0.75 y 0.9 indicaron una buena fiabilidad, y los valores superiores a 0.90 indicaron una excelente fiabilidad. La interpretación del CCI se basó en los límites inferior y superior de los intervalos de confianza del 95%.

Todos los análisis se realizaron con el programa Statistical Package for the Social Sciences (SPSS, versión 23, IBM Corporation, Chicago, IL), y se fijó el nivel de significatividad estadística en $p < 0.05$.

3. Resultados

3.1 Fiabilidad inter-observador de las habilidades locomotrices

En la Tabla 3 se muestra la fiabilidad inter-observador entre los tres evaluadores (CCI e intervalo de confianza del 95%) de las habilidades locomotrices. Las habilidades de carrera y salto con un pie tuvieron una baja fiabilidad inter-observador, tanto inicialmente como tras el proceso de consenso de los evaluadores sobre los criterios de rendimiento. En el caso de las demás habilidades locomotrices, la fiabilidad inter-observador fue mayor después el proceso de elaboración del consenso en todas las comparaciones por pares entre evaluadores. En la primera fase, la fiabilidad inter-observador calculada a través de las mediciones de los tres evaluadores fue al menos moderada en dos habilidades (galope y skip: de moderada a buena en ambas habilidades); sin embargo, en las segunda evaluación tras el consenso, fue de buena a excelente para el desplazamiento lateral y el skip y de moderada a excelente para el galope y el salto horizontal. Por lo tanto, la fiabilidad de las puntuaciones de estos tres evaluadores mejoró tras el consenso entre evaluadores.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.2 Fiabilidad inter-observador de las habilidades de control de objetos

En la Tabla 4 se muestran la fiabilidad inter-observador (CCI e intervalo de confianza del 95%) de las habilidades de control de objetos. La fiabilidad inter-observador aumentó tras la elaboración del consenso sobre los criterios de rendimiento del TGMD-3. Hubo una baja fiabilidad inter-observador para las habilidades de patada y lanzamiento por debajo en las evaluaciones iniciales (primera fase; CCI < 0.5 en todas las comparaciones), pero los valores de CCI asociados a estas dos habilidades superaron el umbral de 0.50 en todos los análisis después del consenso alcanzado entre los evaluadores**.** Una mejora particularmente evidente ocurrió en el caso del lanzamiento por debajo, para el cual la fiabilidad inter-observador después del consenso fue de moderada a excelente en todas las comparaciones por pares. Además, los evaluadores alcanzaron una fiabilidad de moderada a excelente en tres habilidades (bateo, golpeo con raqueta y lanzamiento por debajo) y una fiabilidad de buena a excelente en una habilidad (lanzamiento por encima) tras la elaboración de consenso, mientras que, inicialmente, sólo 1 de las 7 habilidades de control de objetos (bateo) alcanzó una fiabilidad moderada.

3.3 Fiabilidad inter-observador de las puntuaciones de las subescalas del TGMD-3 y de la puntuación global

En la Tabla 5 se muestra la fiabilidad inter-observador (CCI e intervalo de confianza del 95%) de las puntuaciones de las subescalas y de la puntuación global. Los evaluadores mostraron mejoras al comparar las evaluaciones iniciales con las evaluaciones nueve meses después y tras la elaboración del consenso. En el caso de la subescala locomotriz, hubo una mejora de la fiabilidad en todos los análisis, es decir, una fiabilidad de pobre a buena inicialmente pasó a una fiabilidad de moderada a buena tras el consenso. En lo que respecta a la subescala de control de objetos, la fiabilidad entre los evaluadores A y B fue peor (de moderada a excelente) antes del consenso en comparación con después del mismo (de buena a excelente). La fiabilidad de la puntuación global permaneció estable en ambas evaluaciones (primera y segunda fase).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tabla 3. Fiabilidad inter-observador*. Subescalas y global. [Elaboración propia]

Evaluadores	Test	Carrera	Desplazamiento lateral	Galope	Salto a un pie	Salto horizontal	Skip	CCI y límite inferior ≥ 0.5
Evaluador AxB	No-c	0.090 (-0.309-0.462)	0.437 (0.058-0.705)	0.687 (0.407-0.849)	0.251 (-0.152-0.583)	0.688 (0.409-0.849)	**0.805 (0.607-0.909)**	1/6
	Co	0.209 (-0.796-0.651)	**0.968 (0.927-0.986)**	**0.922 (0.823-0.966)**	**0.877 (0.720-0.946)**	**0.941 (0.867-0.974)**	**0.985 (0.967-0.994)**	5/6
Evaluador AxC	No-c	-0.106 (-0.474-0.295)	**0.749 (0.509-0.881)**	**0.771 (0.546-0.892)**	0.470 (0.100-0.726)	0.456 (0.082-0.717)	0.735 (0.485-0.874)	2/6
	Co	-0.246 (-0.579-0.157)	**0.958 (0.905-0.982)**	**0.877 (0.722-0.946)**	0.393 (-0.378-0.732)	**0.877 (0.720-0.946)**	**0.904 (0.781-0.957)**	4/6
Evaluador BxC	No-c	0.683 (0.401-0.847)	0.640 (0.336-0.824)	0.704 (0.434-0.858)	0.423 (0.042-0.697)	0.591 (0.263-0.796)	0.655 (0.358-0.832)	0/6
	Co	0.376 (-0.416-0.725)	**0.922 (0.822-0.965)**	**0.878 (0.724-0.946)**	0.493 (-0.150-0.777)	**0.913 (0.801-0.961)**	**0.939 (0.861-0.973)**	4/6
Todos evaluadores	No-c	0.294 (0.050-0.556)	0.647 (0.439-0.810)	**0.721 (0.539-0.854)**	0.413 (0.167-0.650)	0.459 (0.216-0.684)	**0.774 (0.616-0.884)**	2/6
	Co	0.094 (-0.761-0.571)	**0.906 (0.827-0.954)**	**0.807 (0.665-0.902)**	0.485 (0.244-0.703)	**0.836 (0.711-0.918)**	**0.894 (0.807-0.948)**	4/6

*CCI: medidas únicas, consistencia, modelo de efectos aleatorios de bidireccional, intervalo de confianza del 95%.
No-c: Antes del consenso de los evaluadores; Co: Después del consenso de los evaluadores; en gris, habilidades con CCI ≥ 0.5; en gris y negrita, habilidades con al menos una fiabilidad inter-observador moderada (CCI y límite inferior ≥ 0.5).

Tabla 4. Fiabilidad inter-observador*. Control de objetos. [Elaboración propia]

Evaluadores	Test	Bateo	Bote estático	Lanzamiento por encima	Patada	Golpeo con raqueta	Recepción	Lanzamiento por debajo	CCI y límite inferior ≥ 0.5
Evaluador AxB	No-c	**0.788 (0.576-0.901)**	0.631 (0.322-0.819)	0.660 (0.365-0.834)	0.478 (0.110-0.730)	0.674 (0.387-0.842)	**0.929 (0.845-0.968)**	0.475 (0.106-0.729)	2/7
	Co	**0.972 (0.937-0.987)**	**0.900 (0.772-0.956)**	**0.920 (0.819-0.965)**	0.730 (0.387-0.881)	**0.932 (0.846-0.970)**	**0.852 (0.693-0.932)**	**0.799 (0.545-0.912)**	6/7
Evaluador AxC	No-c	0.732 (0.480-0.872)	0.545 (0.199-0.770)	0.476 (0.107-0.729)	0.412 (0.028-0.690)	0.378 (-0.012-0.668)	0.297 (-0.103-0.614)	0.322 (-0.076-0.631)	0/7
	Co	**0.860 (0.683-0.938)**	0.393 (-0.377-0.733)	0.766 (0.470-0.897)	**0.817 (0.584-0.919)**	**0.884 (0.736-0.949)**	0.554 (-0.012-0.803)	**0.817 (0.584-0.919)**	4/7
Evaluador BxC	No-c	0.615 (0.298-0.810)	0.477 (0.109-0.730)	0.570 (0.234-0.785)	0.378 (-0.012-0.668)	0.409 (0.024-0.688)	0.222 (-0.181-0.562)	0.377 (-0.014-0.667)	0/7
	Co	**0.882 (0.732-0.948)**	0.574 (0.034-0.812)	0.754 (0.441-0.891)	0.718 (0.361-0.876)	**0.882 (0.752-0.946)**	0.660 (0.229-0.850)	**0.797 (0.540-0.911)**	3/7
Todos evaluadores	No-c	**0.711 (0.524-0.848)**	0.551 (0.320-0.748)	0.581 (0.355-0.767)	0.413 (0.166-0.650)	0.483 (0.242-0.701)	0.552 (0.321-0.749)	0.397 (0.150-0.638)	1/7
	Co	**0.848 (0.731-0.924)**	0.480 (0.238-0.699)	**0.874 (0.756-0.941)**	0.612 (0.394-0.788)	**0.851 (0.735-0.926)**	0.602 (0.381-0.781)	**0.860 (0.727-0.934)**	4/7

*CCI: medidas únicas, consistencia, modelo de efectos aleatorios bidireccional, intervalo de confianza del 95%;
No-c: Antes del consenso de los evaluadores; Co: Después del consenso de los evaluadores; en gris, habilidades con CCI ≥ 0.5; en gris y negrita, habilidades con al menos una fiabilidad inter-observador moderada (CCI y límite inferior ≥ 0.5).

Tabla 5. Fiabilidad inter-observador*. Subescalas y global. [Elaboración propia]

Evaluadores	Test	Locomotriz	Control de objetos	Global	CCI y límite inferior ≥ 0.5
Evaluador AxB	No-c	0.499 (0.137-0.743)	**0.787 (0.574-0.900)**	**0.820 (0.634-0.917)**	2/3
	Co	**0.912 (0.811-0.960)**	**0.912 (0.811-0.960)**	**0.945 (0.880-0.976)**	3/3
Evaluador AxC	No-c	0.590 (0.261-0.796)	0.671 (0.382-0.840)	**0.812 (0.618-0.912)**	1/3
	Co	**0.791 (0.582-0.902)**	0.706 (0.438-0.859)	**0.797 (0.592-0.905)**	2/3
Evaluador BxC	No-c	0.621 (0.307-0.813)	0.618 (0.303-0.812)	**0.813 (0.621-0.913)**	1/3
	Co	**0.758 (0.525-0.886)**	0.693 (0.416-0.852)	**0.766 (0.538-0.890)**	2/3
Todos evaluadores	No-c	0.569 (0.341-0.760)	**0.695 (0.502-0.838)**	**0.815 (0.678-0.907)**	2/3
	Co	**0.824 (0.692-0.911)**	**0.778 (0.621-0.886)**	**0.842 (0.721-0.921)**	3/3

*CCI: medidas únicas, consistencia, modelo de efectos aleatorios bidireccional, intervalo de confianza del 95%. No-c: Antes del consenso de los evaluadores; Co: Después del consenso de los evaluadores; en gris, habilidades con CCI ≥ 0.5; en gris y negrita, habilidades con al menos una fiabilidad inter-observador moderada (CCI y límite inferior ≥ 0.5).

Discusión

El objetivo de este estudio fue determinar la fiabilidad inter-observador de las habilidades motrices básicas que componen el TGMD-3 antes y después de la elaboración de un consenso entre los evaluadores sobre los criterios de rendimiento de cada habilidad. La fiabilidad inter-observador mejoró tras el consenso entre los evaluadores en las habilidades locomotrices y las habilidades de control de objetos del TGMD-3 por separado, y en las puntuaciones de las subescalas locomotrices y de control de objetos.

Más concretamente, en el caso de las habilidades locomotrices, los evaluadores alcanzaron una fiabilidad inter-observador de al menos moderada al evaluar 4 de las 6 habilidades tras el consenso, en comparación la fiabilidad inter-observador alcanzada sólo en 2 de las 6 habilidades en la evaluación inicial. Después de la elaboración del consenso, las evaluaciones que más mejoraron fueron las de las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

habilidades locomotrices de desplazamiento lateral y salto horizontal, pasando de una fiabilidad de pobre a buena (desplazamiento lateral) y de pobre a moderada (salto horizontal) inicialmente a una fiabilidad de buena a excelente (desplazamiento lateral) y de moderada a excelente (salto horizontal) tras el consenso. Esta mejora significativa podría demostrar la importancia y el valor de que los evaluadores objetiven colectivamente los criterios de rendimiento. Las habilidades locomotrices con menor fiabilidad fueron la carrera y el salto con un pie en ambas fases, en consonancia a investigaciones previas que también han documentado una menor fiabilidad inter-observador para estas habilidades (14,17,18), posiblemente debido a la subjetividad o complejidad adicional en sus criterios de rendimiento. Por ejemplo, uno de los criterios en la carrera es "*el contacto del pie se reduce al talón o punta (no con el pie plano)*", que es difícil de percibir en cada zancada en las evaluaciones en directo o incluso en las evaluaciones en vídeo, ya que la cámara debe estar lo suficientemente lejos como para grabar 20 metros de carrera. Además, diferentes estudios previos han informado de que la habilidad de salto con un pie es la que presenta una menor fiabilidad inter-observador en el TGMD-3 (14, 18, 19). En contraposición, Maeng et al. (17) encontraron la fiabilidad inter-observador más alta en el salto con un pie. En esta habilidad, dos criterios de rendimiento podrían ser particularmente subjetivos: "*La pierna que no salta se balancea hacia delante de forma pendular para producir fuerza*" y "*Los brazos se flexionan y se balancean hacia delante para producir fuerza*". Especialmente la parte de estos criterios que describe "*... para producir fuerza*" requiere que los evaluadores diferencien entre el balanceo como un movimiento natural de equilibrio y el balanceo específicamente para producir fuerza.

Investigaciones anteriores también han demostrado la complejidad de la evaluación de las habilidades de control de objetos (11) y, como consecuencia, estas habilidades han mostrado una menor fiabilidad inter-observador que las habilidades locomotrices (14). Sin embargo,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

en nuestro estudio, incluso en estas habilidades más complejas, los evaluadores mejoraron su fiabilidad inter-observador desde sus evaluaciones iniciales hasta después del proceso de elaboración del consenso. En este sentido, el acuerdo entre los tres evaluadores fue al menos moderada en sólo 1 habilidad de control de objetos antes del consenso, alcanzando al menos una fiabilidad inter-observador de moderada a excelente en 4 habilidades en la segunda etapa del estudio.

En el presente estudio, las habilidades de control de objetos más problemáticas fueron el bote estático, la recepción y la patada. En base a la literatura existente, no se puede establecer un acuerdo sobre la calidad de la fiabilidad entre evaluadores para estas habilidades. Por ejemplo, Barnet et al. (11) obtuvieron valores de fiabilidad más bajos en la habilidad de recepción, mientras que el acuerdo de los evaluadores para la patada fue excelente. En consonancia con nuestros hallazgos, otros estudios (14, 17, 19, 20) encontraron una fiabilidad entre pobre y moderada para la patada. Una vez más, consideramos que esto podría deberse a la subjetividad de sus criterios de rendimiento: "*Pie de apoyo colocado cerca del balón*". ¿Qué se considera "cerca del balón"? En nuestro estudio, algunos evaluadores sólo consideraban "cerca del balón" si el pie del niño estaba justo al lado del balón, mientras que otros permitían cierta distancia entre el balón y el pie. Estas interpretaciones individuales podrían contribuir a la variación de las puntuaciones y a las fluctuaciones de los índices de fiabilidad entre evaluadores.

La misma discordancia se aplica al bote estático, en el que nuestros resultados indicaron una fiabilidad inter-observador pobre en ambas fases. Sin embargo, la concordancia de los evaluadores fue excelente en la investigación de Maeng et al. (17). En esta habilidad, nuestros evaluadores expresaron la mayor dificultad con los siguientes criterios "*empuja el balón con las yemas de los dedos (sin golpear el balón)*" y "*contacta el balón con una mano a la altura de la cintura aproximadamente*". A pesar de realizar evaluaciones en vídeo y de poder detener la acción y observarla a cámara lenta, nuestros

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

evaluadores encontraron que estos movimientos de corta duración eran difíciles de codificar. En este caso, nuestro índice de fiabilidad inter-observador bajo, incluso después de que los evaluadores tuvieran la oportunidad de llegar a un consenso sobre los criterios de rendimiento, podría deberse a una aplicación incorrecta de uno de sus acuerdos generales: "*Si el evaluador tiene dudas en la valoración de un criterio de rendimiento, lo puntuará como "1"*. Este acuerdo general se estableció teniendo en cuenta que las HMB forman parte del desarrollo motor grueso de los niños, en los que la evaluación del movimiento de un criterio de rendimiento específico no debería ser tan exigente como en la evaluación de un movimiento técnico de, por ejemplo, una actuación deportiva.

En lo que respecta a las puntuaciones de las subescalas, nuestros resultados mostraron una mayor fiabilidad inter-observador en estos índices que en las habilidades individuales. La mayoría de los estudios anteriores también obtuvieron una fiabilidad inter-observador de buena a excelente en las puntuaciones de las subescalas locomotriz y de control de objetos y en la puntuación global (18,21-27), y este nivel de fiabilidad inter-observador fue mayor que en nuestro estudio. En cualquier caso, la fiabilidad relacionada con las puntuaciones de las subescalas en nuestro estudio aumentó después del consenso, especialmente en las puntuaciones de la subescala locomotriz que mejoraron de una fiabilidad inter-observador de pobre a buena a una de moderada a excelente.

Nuestros resultados diferentes a los de estudios anteriores pueden deberse a que utilizamos descriptores más exigentes para nuestros valores de CCI (16). Por ejemplo, solo consideramos los valores de CCI superiores a 0.75 como buenos y los superiores a 0.90 como de excelente fiabilidad. En cambio, desde una perspectiva cualitativa, otros estudios consideraron los valores de CCI superiores a 0.6 como buenos y los superiores a 0.75 como excelentes (12, 19, 21, 24, 28). De hecho, si hubiéramos interpretado el CCI igual que estos estudios anteriores, nuestra fiabilidad inter-observador también habría sido de

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

buena a excelente en las subescalas y en las puntuación global del TGMD-3. Además, aunque algunos estudios solo utilizaron los valores del CCI para interpretar la fiabilidad (11, 12, 17-22, 29, 30), nosotros seguimos las recomendaciones de Koo y Li (16) y utilizamos tanto los límites inferiores como los superiores del intervalo de confianza del 95%. Esto, junto con el hecho de que varios investigadores realizaron otras pruebas estadísticas para evaluar la fiabilidad, como el kappa (23) o el coeficiente de correlación de Pearson (25), podría explicar por qué la fiabilidad inter-observador fue menor en el presente estudio a pesar de que nuestros evaluadores mejoraron tras el consenso. Aunque en investigaciones anteriores se ha señalado la necesidad de llegar a un consenso entre los evaluadores antes de la evaluación (11-13, 17), nuestro estudio es el primero que muestra los beneficios de seguir esta recomendación.

Limitaciones y directrices para futuras investigaciones

Hay que señalar algunas limitaciones de este estudio. En primer lugar, se utilizó una pequeña muestra de 25 niños sanos, y nuestros resultados deben ser replicados con una muestra más grande y más diversa (p. ej.: diferentes orígenes socioeconómicos y preferencias de lateralidad) y con un mayor número de evaluadores con diferentes experiencias (p. ej.: expertos y no expertos en la evaluación de HMB, maestros y profesionales de la salud). Esto permitiría una comprensión más precisa del acuerdo de los evaluadores para las mediciones de la competencia motriz con baterías de pruebas orientadas al proceso. También a tener en cuenta que los 25 participantes fueron evaluados en dos ocasiones. Dado que Ulrich (7) sugirió que los evaluadores pueden mantener el sesgo de memoria después de un período de lavado de 14 días, introdujimos un intervalo de 9 meses en el presente estudio para evitar este posible factor de confusión.

En lo que respecta a las mejoras generales en la metodología que se necesitan en esta línea de investigación, el uso de diferentes análisis estadísticos entre los estudios complica la comparación de los

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

resultados de fiabilidad, y sería valioso estandarizar este enfoque. Aún más importante es que en lugar de crear un consenso entre los evaluadores sobre los criterios de rendimiento en estudios separados, es necesario llegar a un acuerdo común entre todos los usuarios del TGMD-3. De esta forma, la fiabilidad inter-observador no sólo mejoraría dentro de un mismo equipo de investigación, sino entre otros equipos y profesionales clínicos.

Conclusiones

Este estudio muestra que el acuerdo de los evaluadores sobre los criterios de rendimiento podría mejorar la fiabilidad inter-observador en la evaluación de las habilidades del TGMD-3. Recomendamos que tanto en el ámbito de investigación como en el ámbito clínico cuando se use el TGMD-3 para guiar las intervenciones curriculares apropiadas en Educación Física, se preste atención a la elaboración de un consenso sobre los criterios de rendimiento complejos para mejorar el acuerdo entre evaluadores. Lo ideal sería que se compartieran los criterios de rendimiento mejorados entre los estudios para permitir un consenso común y una formación específica y estandarizada para todos los usuarios del TGMD-3.

Referencias

1. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. J Sci Med Sport 2019;22(3):311–8.
2. Scheuer C, Herrmann C, Bund A. Motor tests for primary school aged children: A systematic review. J Sports Sci 2019;37(10):1097–112.
3. Tamplain P, Webster EK, Brian A, Valentini NC. Assessment of motor development in childhood: Contemporary issues, considerations, and future directions. J Mot Learn Dev. 2020;8(2):391–409.
4. Eddy LH, Bingham DD, Crossley KL, Shahid NF, Ellingham-Khan M, Otteslev A, et al. The validity and reliability of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. PLoS One 2020;15(8):e0237919.
5. Stodden DF, Goodway JD, Langendorfer SJ, Roberton MA, Rudisill ME, Garcia C, et al. A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. Quest 2008;60(2):290–306.
6. Klingberg B, Schranz N, Barnett LM, Booth V, Ferrar K. The feasibility of fundamental movement skill assessments for pre-school aged children. J Sports Sci 2019;37(4):378–86.
7. Ulrich DA. Test of Gross Motor Development - 3 (Third Ed). Pro-Ed Publishers; 2019.
8. Rey E, Carballo-Fazanes A, Varela-Casal C, Abelairas-Gómez C, ALFA-MOV Project collaborators. Reliability of the test of gross motor development: A systematic review. PLoS One 2020;15(7):e0236070.
9. Devellis RF. Scale Development: Theory and Application. Thousand Oaks, CA: Sage Publications; 2012.
10. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use. Cary, NC, Estados Unidos de América: Oxford University Press; 2014.
11. Barnett LM, Minto C, Lander N, Hardy LL. Interrater reliability assessment using the Test of Gross Motor Development-2. J Sci Med Sport 2014;17(6):667–70.
12. Houwen S, Hartman E, Jonker L, Visscher C. Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. Adapt Phys Activ Q 2010;27(2):143–59.
13. Cano-Cappellacci M, Leyton FA, Carreño JD. Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica 2015;49(0).
14. Carballo-Fazanes A, Rey E, Valentini NC, Rodríguez-Fernández JE, Varela-Casal C, Rico-Díaz J, et al. Intra-Rater (Live vs . Video Assessment) and Inter-Rater (Expert vs . Novice) Reliability of the Test of Gross Motor Development - Third

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Edition. International Journal of Environmental Research and Public Health. 2021;18: 1652.
15. Webster K. TGMD-3 Administration [Internet]. Youtube; 2014 [citado el 3 de agosro de 2022]. Disponible en: https://www.youtube.com/watch?v=9WggHyZpXl0.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63.
17. Maeng H, Webster EK, Pitchford EA, Ulrich DA. Inter- and intrarater reliabilities of the Test of Gross Motor Development—third edition among experienced TGMD-2 raters. Adapt Phys Activ Q 2017;34(4):442–55.
18. Valentini NC, Zanella LW, Webster EK. Test of Gross Motor Development—third edition: Establishing content and construct validity for Brazilian children. J Mot Learn Dev 2017;5(1):15–28.
19. Rintala PO, Sääkslahti AK, Iivonen S. Reliability assessment of scores from video-recorded TGMD-3 performances. J Mot Learn Dev 2017;5(1):59–68.
20. Kim Y, Park I, Kang M. Examining rater effects of the TGMD-2 on children with intellectual disability. Adapt Phys Activ Q 2012;29(4):346–65.
21. Estevan I, Molina-García J, Queralt A, Álvarez O, Castillo I, Barnett L. Validity and reliability of the Spanish version of the Test of Gross Motor Development–3. J Mot Learn Dev 2017;5(1):69–81.
22. Kim S, Kim MJ, Valentini NC, Clark JE. Validity and reliability of the TGMD-2 for South Korean children. J Mot Behav 2014;46(5):351–6.
23. Lopes VP, Saraiva L, Rodrigues LP. Reliability and construct validity of the test of gross motor development-2 in Portuguese children. Int J Sport Exerc Psychol 2018;16(3):250–60.
24. Mohammadi F, Bahram A, Khalaji H, Ulrich DA, Ghadiri F. Evaluation of the psychometric properties of the Persian version of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the test of Gross Motor Development–3rd Edition. J Mot Learn Dev 2019;7(1):106–21.
25. Simons J, Daly D, Theodorou F, Caron C, Simons J, Andoniadou E. Validity and reliability of the TGMD-2 in 7-10-year-old Flemish children with intellectual disability. Adapt Phys Activ Q 2008;25(1):71–82.
26. Valentini NC. Validity and reliability of the TGMD-2 for Brazilian children. J Mot Behav 2012;44(4):275–80.
27. Wagner MO, Webster EK, Ulrich DA. Psychometric properties of the test of Gross Motor Development, third Edition (German translation): Results of a pilot study. J Mot Learn Dev 2017;5(1):29–44.
28. Capio CM, Eguia KF, Simons J. Test of gross motor development-2 for Filipino children with intellectual disability: validity and reliability. J Sports Sci 2016;34(1):10–7.
29. Ayán C, Cancela J, Sánchez-Lastra M, Carballo-Roales A, Domínguez-Meis F, Redondo-Gutiérrez L. Fiabilidad y Validez de la Batería TGMD-2 en Población Española. Rev Iberoam Diagn Eval - Aval Psicol 2019;51(1):21-33.
30. Aye, T., Oo, K. S., Khin, M. T., Kuramoto-Ahuja, T., & Maruyama, H. (2017). Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. *Journal of Physical Therapy Science*, *29*(10), 1726–1731. https://doi.org/10.1589/jpts.29.1726.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.2 *ALFAMOV*: LA HERRAMIENTA PARA LA EVALUACIÓN DE LAS HABILIDADES MOTRICES BÁSICAS

Alfamov es la primera herramienta digital en formato *app* creada para la evaluación y el seguimiento sistemático de la competencia en las HMB de niños y niñas en la etapa primaria (de los 6 a los 12 años). Fue diseñada y desarrollada dentro del proyecto [RTI2018-096106-A-I00] financiado por FEDER/ Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación.

La creación y diseño de *Alfamov* se realizó con un equipo multidisciplinar de investigadores de la Universidade de Santiago de Compostela y la Universidade de Vigo. El equipo se formó con profesionales del ámbito de la educación física, del diseño gráfico y audiovisual y del ámbito sanitario (pediatría y enfermería). Paralelamente, se contó con la colaboración de una empresa informática que se encargó del desarrollo del software a partir de la información aportada por el equipo investigador de *Alfamov*.

Las pruebas científicas presentadas en el apartado 6.1 de resultados han servido como fundamento para la creación de la herramienta digital *Alfamov*. Por tanto, los resultados de fiabilidad obtenidos en los estudios se consideraron requisitos que debía cumplir *Alfamov* para ser utilizada en la práctica clínica. Tomando como base estos hallazgos, el proceso de desarrollo de *Alfamov* consistió en el trabajo coordinado del equipo investigador. Los profesionales de la salud y de la educación física decidieron sobre el contenido que debería incluir la herramienta y la profesional de diseño gráfico y audiovisual determinó cómo este contenido debería ser organizado de forma visual.

Más concretamente, nuestro equipo se encargó de diseñar la retícula compositiva de base para estructurar la articulación de los contenidos visuales de la *app*. En la figura 7 se puede observar el esquema dibujado a mano alzada del primer diseño de navegación. Por otro lado, nuestro equipo también se encargó de definir los elementos base de identidad visual de *Alfamov*: a) la marca corporativa (Figura 8), b) la tipografía (Figura 9), c) la gama de color (Figura 10), y d) el diseño de los iconos contenidos en la *app* (Figura 11). Igualmente fue responsable de la dirección de arte y realización de todas las imágenes fotográficas y secuencias de vídeo incluidas en la *app*.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

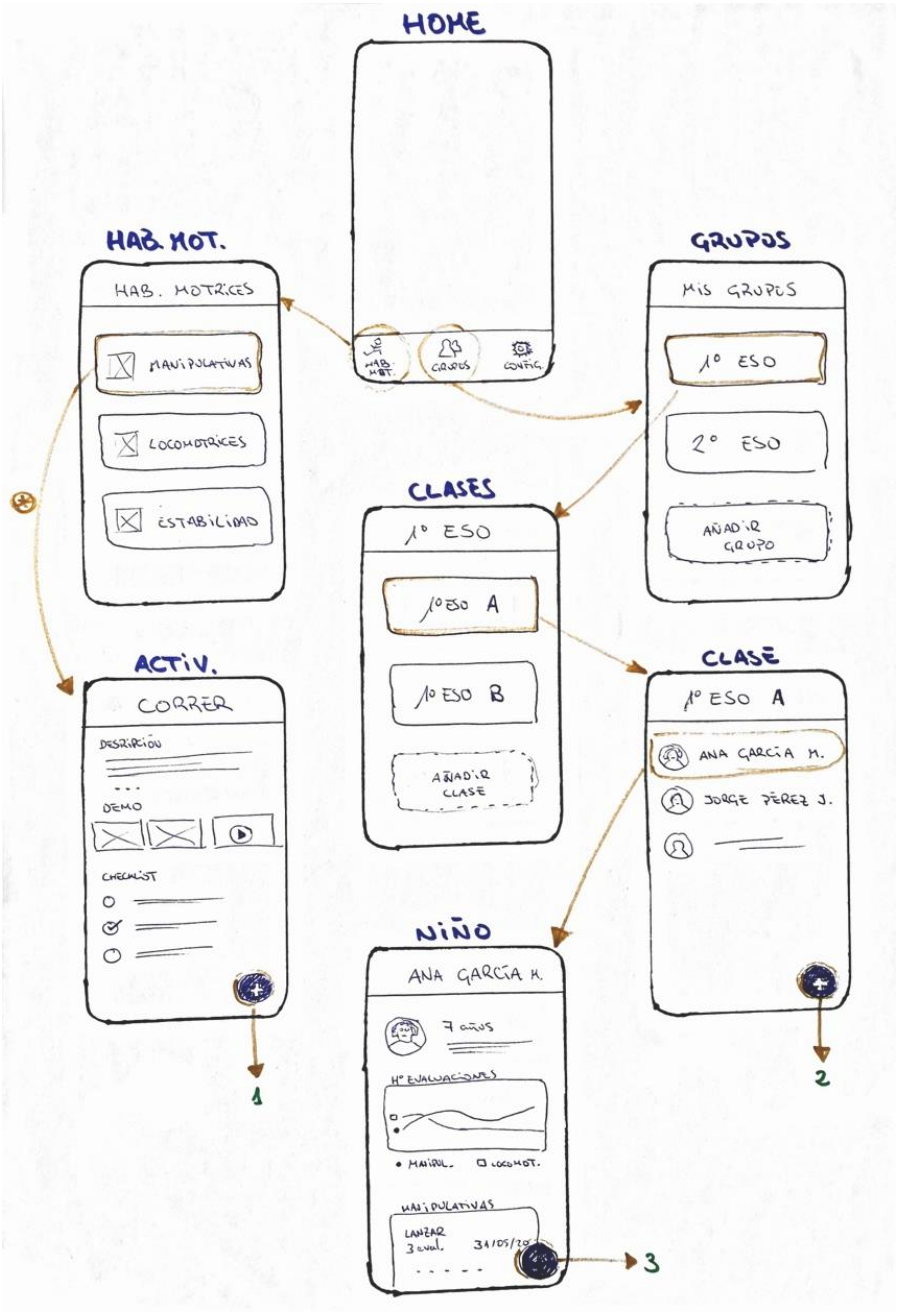


Figura 7. Esquema a mano alzada del diseño de navegación. [Elaboración propia]



Figura 8. Marca corporativa de *Alfamov*. [Elaboración propia]

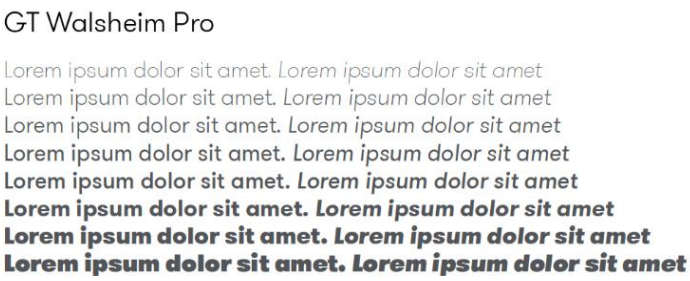


Figura 9. Tipografía seleccionada para *Alfamov*. [Elaboración propia]



Figura 10. Gama de colores utilizados en *Alfamov*. [Elaboración propia]

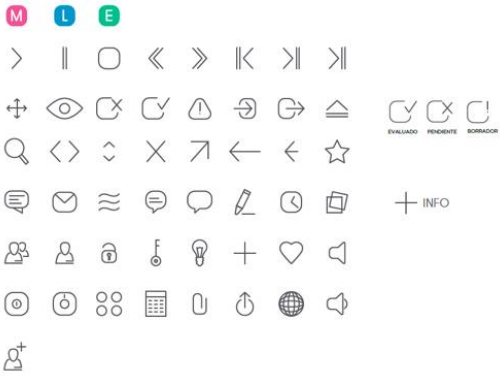


Figura 11. Diseño de los iconos de *Alfamov*. [Elaboración propia]

A partir del esquema inicial de navegación y del diseño del interfaz, se inició el trabajo coordinado con la empresa informática desarrolladora de software. Mediante procesos de *feedback* constantes se fue desarrollando la *app* hasta su versión final. En la Figura 12 se puede observar la evolución del diseño, y en la Figura 13 un ejemplo del proceso de *feedback* entre nuestro equipo de investigación y la empresa informática.

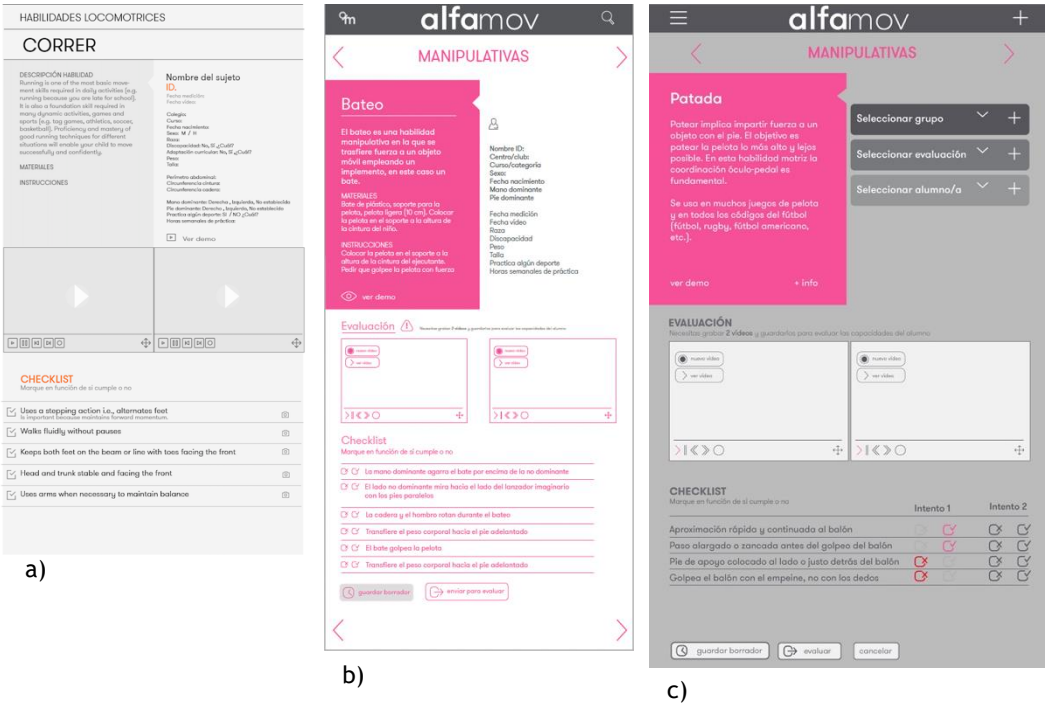


Figura 12. Evolución del diseño visual: versión inicial (a), versión intermedia (b), y versión final (c). [Elaboración propia]

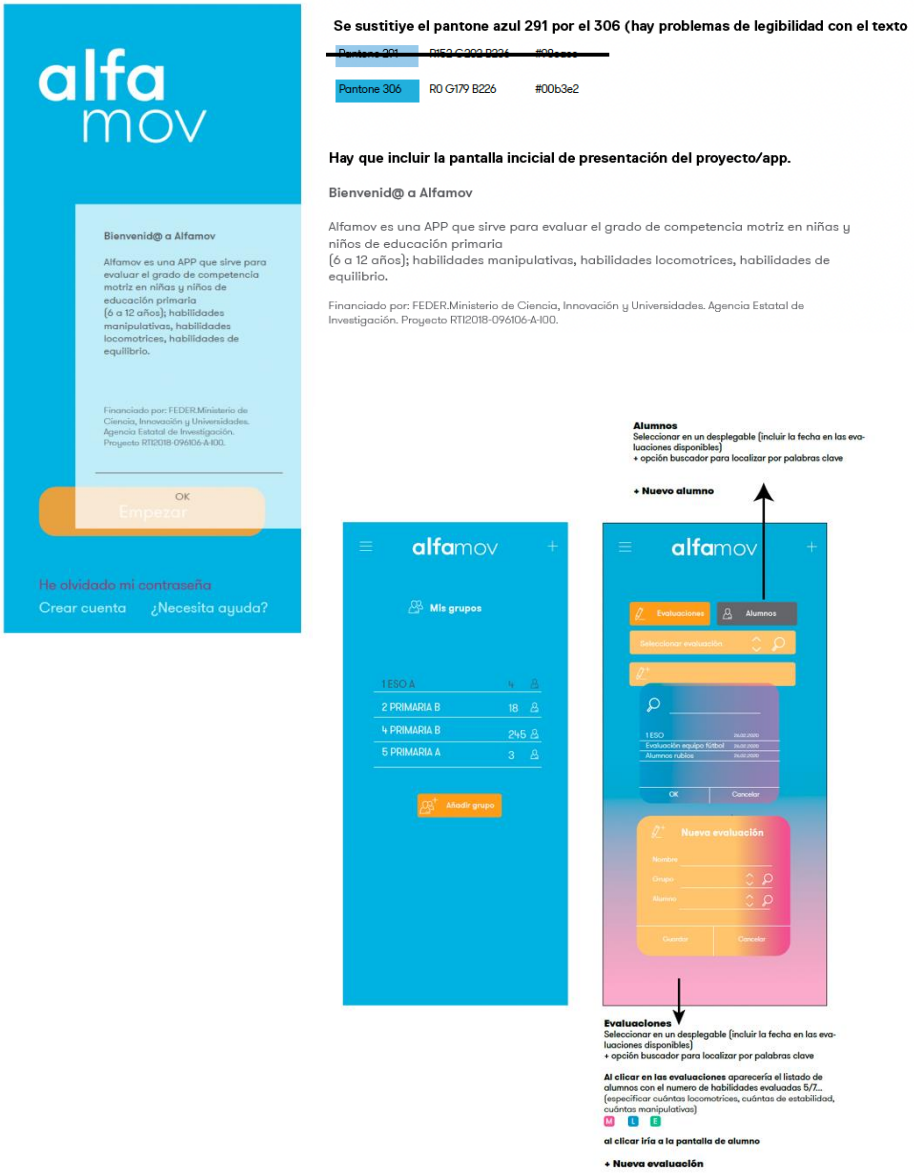


Figura 13. Comunicación para aplicar cambios entre el equipo de investigación de *Alfamov* y la empresa desarrolladora de software. [Elaboración propia]

La versión final de *Alfamov* (Figura 14) incluye 22 HMB [10 habilidades locomotrices, 7 manipulativas y 5 de estabilidad (Tabla 5)] seleccionadas de diferentes herramientas de evaluación[80,117,119,128,134] de las HMB tras una exhaustiva revisión de la literatura científica. *Alfamov*, cuenta con dos áreas diferenciadas:

a) El área didáctica, con información sobre las HMB en general y su potencial importancia [Figura 14, apartados c) y d)] así como información específica de cada una de las habilidades incluidas y también un vídeo demo de la ejecución correcta de la habilidad [apartado e) Figura 14].
b) El área de evaluación de las HMB, que puede ser realizada a tiempo real (grabando a un niño o niña realizando las habilidades) o en diferido (eligiendo un vídeo grabado previamente desde la galería de imágenes) [apartado e) Figura 14].

Tabla 5. Habilidades motrices básicas incluidas en *AlfaMov*. [Elaboración propia]

Locomotrices	*Manipulativas*	*Estabilidad*
Carrera[1]	Recepción[1]	Equilibrio estático[2]
Salto con un pie[1]	Patada[1]	Equilibrio dinámico[3]
Salto vertical[2]	Bote estático[1]	Voltereta hacia delante[3]
Galope[1]	Golpeo con raqueta[1]	Voltereta hacia atrás[4]
Skip[1]	Bateo[1]	Rodar[4]
Trepar[3]	Lanzamiento por encima[1]	
Zancada[1]	Lanzamiento por debajo[1]	
Salto horizontal[1]		
Cambio de dirección[2]		
Desplazamiento lateral[1]		

[1] Habilidades extraídas del Test of Gross Motor Development[119,128]
[2] Habilidades extraídas de Get Skilled Get Active[117]
[3] Habilidades extraídas de 'Fundamental Movement Skills. The tools for learning, teaching and assessment.'[134]
[4] Habilidades extraídas de 'Fundamental Movement Skills are more than run, throw and catch: the role of stability skills'[80]

Para facilitar la evaluación, *Alfamov* dispone de una lista de comprobación con los criterios de rendimiento de cada habilidad para que sean puntuados por los evaluadores [apartado e) Figura 14]. Estos criterios de rendimiento se puntuarán con "1" si se cumplen o "0" si, por el contrario, no se cumplen. La suma de los criterios de rendimiento de cada habilidad dará lugar a la puntuación de cada habilidad individualmente. Además, *Alfamov* ofrece la puntuación de las

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

subescalas (locomotriz, manipulativa y de estabilidad) sumando los resultados de las habilidades que las componen y, finalmente, ofrece la puntuación global de competencia motriz. En la Tabla 6 se puede observar el número de criterios de rendimiento de cada habilidad, así como la puntuación máxima que se podrá alcanzar en cada una, en las subescalas y en la puntuación global de competencia de las HMB.

Se puede acceder a la información de las puntuaciones tras la evaluación a través del perfil creado para cada niño o niña [apartado f) Figura 14] donde también se podrán registrar las características antropométricas u otros datos de interés (p. ej.: la dominancia de mano o pie) y observar las evaluaciones realizadas con anterioridad, así como detalles de cada evaluación (como la fecha de evaluación o si alguna habilidad está incompleta).

Tabla 6. Habilidades, criterios de rendimiento y puntuaciones de *AlfaMov*.[Elaboración propia]

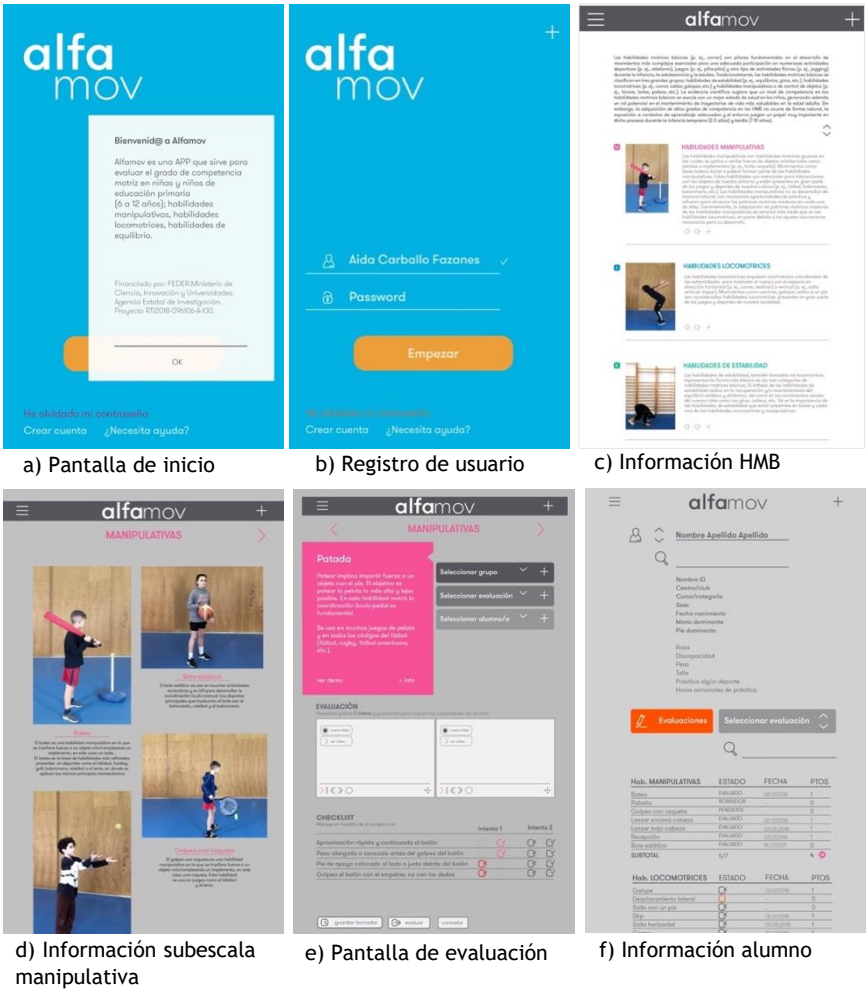
Habilidades locomotrices			Habilidades de control de objetos			Habilidades de estabilidad		
Habilidades	**NC**	**PM**	**Habilidades**	**NC**	**PM**	**Habilidades**	**NC**	**PM**
Carrera	4	8	Recepción	3	6	Equilibrio estático	5	10
Salto con un pie	4	8	Patada	4	8	Equilibrio dinámico	4	8
Salto vertical	6	12	Bote estático	3	6	Voltereta delante	5	12
Galope	4	8	Golpeo raqueta	4	8	Voltereta atrás	4	8
Skip	3	6	Bateo	5	10	Rodar	5	10
Trepar	5	10	Lanz. encima	4	8			
Zancada	3	6	Lanz. debajo	4	8			
Salto horizontal	4	8						
Cambio dirección	5	10						
Desplazamiento lateral	4	8						
Puntuación subescalas		*84*			*54*			*48*
Puntuación global		*186*						

NC número de criterios de rendimiento, *PM* puntuación máxima, *Lanz.* lanzamiento

En la Figura 15 se presenta un ejemplo de evaluación de un niño de 10 años utilizando *Alfamov*. Se puede observar en el perfil del niño las puntuaciones obtenidas en cada HMB de forma individual, las puntuaciones de las subescalas y puntuación global de competencia motriz.

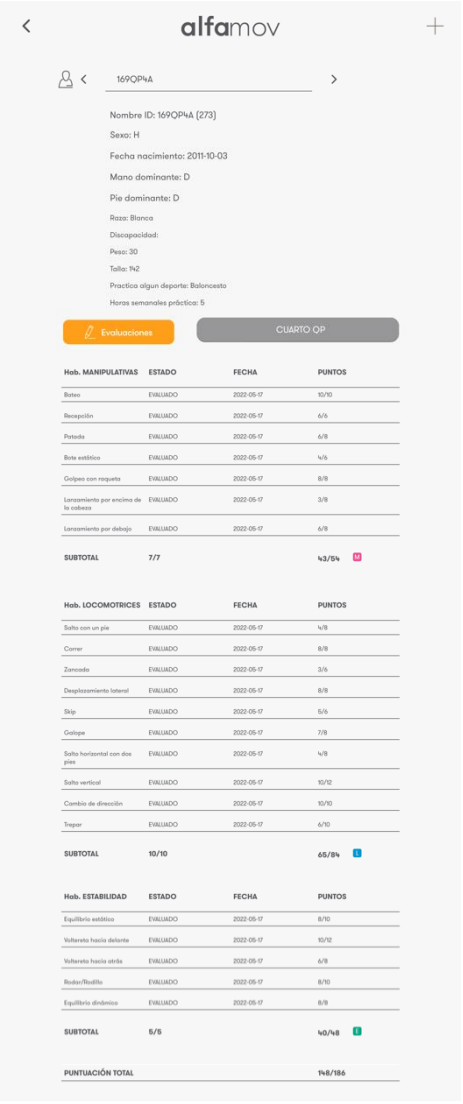
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Para promover la mayor difusión de *Alfamov* se ha creado una página web (https://alfamov.com) con información detallada sobre el presente proyecto, los investigadores que lo conforman y las publicaciones asociadas al mismo. Además, se puede acceder a la descarga de la *app* para Android o iOS (https://alfamov.com/descargar-alfamov/) así como a un videotutorial de uso (**¡Error! Referencia de hipervínculo no válida.**).



a) Pantalla de inicio
b) Registro de usuario
c) Información HMB
d) Información subescala manipulativa
e) Pantalla de evaluación
f) Información alumno

Figura 14. Ejemplos de pantallas del diseño final de *Alfamov*. [Elaboración propia]



Perfil con las características antropométricos y datos de interés del niño.

Puntuaciones obtenidas en cada habilidad categorizadas según a la subescala que pertenecen.

Figura 15. Ejemplo de perfil de un niño tras evaluar su competencia motriz con *Alfamov*. [Elaboración propia]

6.3 PUBLICACIONES DIRECTAMENTE RELACIONADAS CON LA TEMÁTICA DE LA TESIS DOCTORAL

1. Artículo original: **Carballo-Fazanes A**, Rodrigues LP, Silva R, Lopes VP, Abelairas-Gómez C. The Developmental Trajectory of Motor Competence of Children That Lived the COVID-19 Confinement Period: A Four-Year Follow-Up Study in Portuguese Children. J Funct Morphol Kinesiol. 2022; 7(3):64. Doi: 10.3390/jfmk7030064.

2. Artículo original: **Carballo-Fazanes A**, Rodríguez-Fernández JE, Mohedano-Vázquez N, Rodríguez-Núñez A, Abelairas-Gómez C. Competencia motriz y condición física relacionada con la salud en escolares de Educación Primaria (Motor competence and health-related physical fitness in schoolchildren). Retos. 2022; 46:218-26. Doi: 10.47197/retos.v46.93906.

3. Artículo original: Prieto J, Cueto S, **Carballo-Fazanes A**, Abelairas-Gómez C. Psychomotor development disorders in apparently healthy children and considerations of family evaluation. J Human Sport Exer, 2022; 17(3):518-527. Doi: 10.14198/jhse.2022.173.04

4. Artículo original: **Carballo-Fazanes A,** Rico-Díaz J, Barcala-Furelos R, Rey E, Rodríguez-Fernández JE, Varela-Casal C, Abelairas-Gómez C. Physical activity habits and determinants, sedentary behaviour and lifestyle in university students. Int J Environ Res Public Health. 2020;17(9):3272. Doi: 10.3390/ijerph17093272.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.4 OTRAS ACTIVIDADES QUE HAN CONTRIBUIDO A LA FORMACIÓN Y APRENDIZAJE INVESTIGADORA DE LA CANDIDATA

6.4.1 Estancias de investigación durante el período predoctoral

A continuación se muestran las estancias realizadas durante el período predoctoral (un total de 6 meses; Anexo 2).

1. Estancia internacional. Instituto Politécnico de Viana do Castelo (Portugal). Período: 15-03-2022 al 15-06-2022.

   Financiada por el Ministerio de Universidades en la convocatoria: *Orden de 23 de noviembre de 2021 por la que se convocan ayudas complementarias destinadas a beneficiaros de ayudas de Formación del Profesorado Universitario del Subprograma Estatal de Formación en I+D+i*. (Referencia: EST22/00498).

2. Estancia nacional. Universitat de València (España). Período: 01-01-2022 al 01-02-2022.

3. Estancia nacional. Universitat de València (España). Período: 12-04-2021 al 13-06-2021.

   Financiada por el Ministerio de Universidades en la convocatoria: *Orden de 24 de noviembre de 2020 por la que se convocan ayudas complementarias destinadas a beneficiaros de ayudas de Formación del Profesorado Universitario del Subprograma Estatal de Formación en I+D+i*. (Referencia: EST21/00146).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.4.2 Otras publicaciones de la candidata durante el período predoctoral (indexadas en JCR)

Originales

1. Carta al editor: **Carballo-Fazanes A,** Abelairas-Gómez C, Rodríguez-Ruiz E, Barcala-Furelos R, Rodríguez-Núñez A. Anti-choking suction devices use. A pilot simulated study with parents and kindergarten teachers. Resuscitation. 2022;177:5-6. doi: 10.1016/j.resuscitation.2022.06.005.
2. Artículo original: Abelairas-Gómez C, Martinez-Isasi S, Barcala-Furelos R, Varela-Casal C, **Carballo-Fazanes A,** Pichel-López M, et al. Training frequency for educating schoolchildren in basic life support: very brief 4-month rolling-refreshers versus annual retraining-a 2-year prospective longitudinal trial. BMJ Open. 2021;11(11):e052478. Doi: 10.1136/bmjopen-2021-052478.
3. Artículo original: Martínez-Isasi S, Jorge-Soto C, Barcala-Furelos R, Abelairas-Gómez C, **Carballo-Fazanes A**, et al. Performing Simulated Basic Life Support without Seeing: Blind vs. Blindfolded People. Int J Environ Res Public Health. 2021;18(20):10724. Doi: 10.3390/ijerph182010724.
4. Carta al editor: Abelairas-Gómez C, **Carballo-Fazanes A**, López-García S, Rodríguez-Núñez A. Let's train CPR together: mandatory cardiopulmonary resuscitation competencies for undergraduate students in healthcare and education: A step forward to train schoolteachers. Eur J Anaesthesiol. 2021;38(10):1106-1107. Doi: 10.1097/EJA.0000000000001442.
5. Brief Report: **Carballo-Fazanes A**[*], Trastoy-Quintela J, Rey-Noriega C, Sánchez-Rodríguez L, Rodríguez-Núñez A. Paediatric residents deliver similar quality simulated neonatal resuscitation using 3:1 and 15:2 ratios. Acta Paediatr. 2021;110(11):3009-3010. Doi: 10.1111/apa.15991.
6. Artículo original: Abelairas-Gómez C, Schroeder DC, **Carballo-Fazanes A**, Böttiger BW, López-García S,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Martínez-Isasi S, Rodríguez-Núñez A. KIDS SAVE LIVES in schools: cross-sectional survey of schoolteachers. Eur J Pediatr. 2021;180(7):2213-2221. Doi: 10.1007/s00431-021-03971-x.

7. Artículo original: Barcala-Furelos R, Abelairas-Gómez C, Alonso-Calvete A, Cano-Noguera F, **Carballo-Fazanes A**, Martínez-Isasi S, Rodríguez-Núñez A. Safe On-Boat Resuscitation by Lifeguards in COVID-19 Era: A Pilot Study Comparing Three Sets of Personal Protective Equipment. Prehosp Disaster Med. 2021;36(2):163-169. Doi: 10.1017/S1049023X2100011X.

8. Artículo original: **Carballo-Fazanes A,** Barcala-Furelos R, Eiroa-Bermúdez J, Fernández-Méndez M, Abelairas-Gómez C, Martínez-Isasi S, Murciano M, Fernández-Méndez F, Rodríguez-Núñez A. Physiological demands of quality cardiopulmonary resuscitation performed at simulated 3250 meters high. Am J Emerg Med. 2020;38(12):2580-2585. Doi: 10.1016/j.ajem.2019.12.048.

9. Artículo original: **Carballo-Fazanes A**[*], Bierens JJLM, The International Expert Group To Study Drowning Behaviour. The Visible Behaviour of Drowning Persons: A Pilot Observational Study Using Analytic Software and a Nominal Group Technique. Int J Environ Res Public Health. 2020;17(18):6930. doi: 10.3390/ijerph17186930.

10. Artículo original: Abelairas-Gómez C, **Carballo-Fazanes A**, Martínez-Isasi S, López-García S, Rico-Díaz J, Rodríguez-Núñez A. Knowledge and attitudes on first aid and basic life support of Primary and Preschool teachers and parents. An Pediatr (Engl Ed). 2020;92(5):268-276. doi: 10.1016/j.anpedi.2019.10.010.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11. Carta al editor: **Carballo-Fazanes A**[*], Cancela-Patiño P, Neiro-Rey C. ¿Ser corredor aficionado se relaciona con tener buenos hábitos de alimentación? [Being an amateur runner is related to good eating habits?]. Gac Sanit. 2020;34(1):97. Spanish. Doi: 10.1016/j.gaceta.2019.05.017.

12. Carta al editor: Abelairas-Gómez C, **Carballo-Fazanes A**[*], López-García S, Martínez-Isasi S, Rodríguez-Núñez A. School teachers should know how and teach to save lives to the kids. The inclusion of basic life support training in university degrees whose aim is to train teachers. BLS mandatory in school & university. An Pediatr (Engl Ed). 2020;92(5):319-320. Doi: 10.1016/j.anpedi.2019.11.010.

13. Carta al editor: Abelairas-Gómez C, López-García S, Martínez-Isasi S, **Carballo-Fazanes A**[*], Rodríguez-Núñez A. Basic life support knowledge of the future of the Infant and Primary School teacher. An unresolved problem in university study plans? An Pediatr (Barc). 2019;91(5):344-5. Doi: 10.1016/j.anpedi.2018.10.010.

[*]*Autora de correspondencia*

6.4.3 Presentaciones en congresos científicos

En este apartado se presenta la relación de contribuciones a congresos nacionales e internacionales durante el período doctoral que tienen relación con la presente tesis doctoral:

1. Póster: **Carballo-Fazanes A**, Abelairas-Gómez C, Rey E, Varela-Casal C, Rodríguez-Núñez A. "AlfaMov: La aplicación para evaluar de forma sencilla la competencia motriz de escolares: análisis de fiabilidad." *36 Congreso Nacional de la Sociedad Española de Pediatría Extrahospitalaria y Atención Primaria – SEPEAP 2022.* Fecha inicio-fin: 20/10/2022 – 22/10/2022, lugar de celebración: Alicante (España).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2. Comunicación: **Carballo-Fazanes A**, Rodrigues LP, Rey E, Rodríguez-Núñez A, Abelairas-Gómez C. "AlfaMov: Uma app para avaliar a competência motora das crianças." *XVII Seminário de Desenvolvimento Motor da Criança*. Fecha inicio-fin: 28/10/2022 – 29/10/2022, lugar de celebración: Faro, Algarve (Portugal).
3. Comunicación: **Carballo-Fazanes A,** Rodríguez-Núñez A, Rey E, Varela-Casal C, Martínez Isasi S, Abelairas-Gómez C. "Relación entre el desarrollo de las habilidades motrices básicas y la práctica de actividad física organizada en niños de entre 6 y 12 años." *IV Congreso Internacional y VII Nacional de Enfermería y Salud.* Fecha inicio-fin: 26/09/2022 – 27/09/2022, lugar de celebración: León (España).
4. Póster con defensa: **Carballo-Fazanes A**, Abelairas-Gómez C, Rey E, Varela-Casal C, Rodríguez-Núñez A. "AlfaMov: una nueva app para evaluar de forma rápida en la consulta la competencia motriz de los escolares." *35 Congreso Nacional de la Sociedad Española de Pediatría Extrahospitalaria y Atención Primaria.* Fecha inicio-fin: 30/09/2021 – 02/10/2021, lugar de celebración: A Coruña (España).
5. Comunicación: **Carballo-Fazanes A**, Abelairas-Gómez C, Varela-Casal C, Rey E. "Relationship between motor competence and intellectual maturity in Spanish preschool children". *Fifth Assembly of the International Motor Development Research Consortium*. Fecha inicio-fin: 22/09/2022 – 25/09/2022, lugar de celebración: Online.
6. Póster con defensa: Menescardi C, **Carballo-Fazanes A**, Romero-Martínez J, Cuenca R, Estevan I. "Audio-live vs video assessment reliability of motor competence circuit-based field-test" *Fifth Assembly of the International Motor*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Development Research Consortium. Fecha inicio-fin: 22/092022 – 25/09/2022, lugar de celebración: Online.

7. Ponencia en mesa redonda "Xacobeo 2021: Aportando Valor a la Investigación". **Carballo-Fazanes, A**. "La alfabetización motriz como elemento clave para la adquisición de estilos de vida saludables". *III Congreso Internacional y VI Nacional de Enfermería y Salud.* Fecha inicio-fin: 26/04/2021 – 28/04/2021, lugar de celebración: Online.

6.4.4 Premios recibidos:

Por último, se presentan los premios recibidos durante el periodo predoctoral:

1. Premio "Nestlé España, S.A." a la mejor comunicación oral "*¿Podrían futuros profesores de educación primaria resolver un atragantamiento infantil utilizando los nuevos dispositivos LifeVac® y DeCHOKER®*? en el en el 35 Congreso Nacional de la Sociedad Española de Pediatría Extrahospitalaria y Atención Primaria celebrado del 30 de septiembre al 2 de octubre de 2021 en A Coruña.

2. Premio a la mejor comunicación oral *"Nuevos dispositivos de aspiración anti-atragantamiento ¿Serían capaces de utilizarlos los padres y profesores de educación infantil sin recibir una formación previa?"* en el 35 Congreso Nacional de la Sociedad Española de Cuidados Intensivos Pediátricos celebrado en modalidad virtual del 17 al 21 de mayo de 2021.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7 DISCUSIÓN

La adquisición de hábitos saludables entre los que se encuentra la práctica regular de AF es esencial para poder conseguir un estilo de vida saludable. Las enfermedades no transmisibles así como los factores de riesgo como el sobrepeso y obesidad son producto de malos hábitos[7]. No obstante, a pesar de la preocupación a nivel mundial ante el incremento de la inactividad física, sobrepeso y obesidad[5,16] y de la creación de numerosas iniciativas educativas, nutricionales, farmacológicas o de AF, el problema persiste prácticamente sin ningún avance significativo de mejora. Por esta razón, es necesario buscar vías adicionales o complementarias para conseguir el objetivo de una población sana. Teniendo en cuenta que los hábitos saludables se aprenden, interiorizan y adquieren, la infancia es la etapa crucial en la que debemos investigar e intervenir[125].

En este sentido, el foco en la CM como predictor de salud ha permanecido en un segundo plano. No obstante, podría ser muy relevante ya que la competencia en las HMB se ha mostrado como influyente en el desarrollo físico, psicológico, cognitivo y social a lo largo de la vida[35,101,135]. Por ello, es preocupante el deterioro constante de la CM que se está produciendo en los últimos años[86,87]. Para frenar este deterioro es necesario planificar intervenciones específicas, pero antes hay que conocer qué personas se encuentran por debajo de los valores normativos de desarrollo de las HMB.

Sin embargo, la evaluación fiable de las HMB sigue siendo un problema a solventar debido a la falta de tiempo, formación y experiencia por parte de los evaluadores para llevarla a cabo o la necesidad de herramientas apropiadas[76,84,105,136]. Estos inconvenientes están provocando que se descuide la evaluación de las HMB. Teniendo esto en cuenta, la presente tesis doctoral aporta nuevas y necesarias evidencias sobre la evaluación de las HMB.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

En primer lugar, se presenta una revisión sistemática de la fiabilidad del TGMD, uno de los instrumentos más utilizados para la evaluación de las habilidades motrices gruesas de niños y niñas, aportando una síntesis y evaluación de los hallazgos alcanzados por los estudios incluidos en la revisión.

En segundo lugar, se muestran dos estudios observacionales que tratan de aportar evidencia sobre la fiabilidad inter- e intra-observador del TGMD. El primero de ellos, proporciona información sobre la fiabilidad inter- e intra-observador del TGMD entre evaluadores de diferentes perfiles (expertos y novatos) empleando dos métodos de observación (en directo y en vídeo) para evaluar las HMB incluidas en el TGMD. El segundo, partiendo de los menores índices de fiabilidad inter-observador en comparación con los relativos a los de la fiabilidad intra-observador mostrados en la literatura, muestra la mejora de la fiabilidad entre los evaluadores tras un acuerdo previo a la evaluación de cada criterio de rendimiento que conforma las HMB.

Por último, debido a la falta de herramientas digitales para evaluar las HMB que incluya habilidades locomotrices, manipulativas y de estabilidad, esta tesis ha supuesto la generación de un instrumento novedoso, denominado *Alfamov,* que permite la evaluación de la CM de niños y niñas.

7.1 FIABILIDAD DEL TGMD

El TGMD es una de las diversas herramientas orientadas al proceso para evaluar la competencia de las habilidades motrices gruesas mediante la observación en niños y niñas de edad preescolar[104,137] y primaria[137,138]. El TGMD es considerado un test psicométrico robusto, es por ello que continúa siendo una de las primeras opciones a la hora de elegir un instrumento para la evaluación de las HMB en la infancia[138]. La creciente evidencia científica de los últimos años sobre la fiabilidad del TGMD, en sus dos últimas versiones (TGMD-2 y TGMD-3) y en distintos grupos poblacionales (niños con TEA[139], con trastorno por déficit de atención[140], con déficits visuales[141], trastornos mentales y conductuales[142] o discapacidad intelectual[143]) justifica la necesidad de una revisión sistemática, presentada como primer resultado de esta tesis doctoral, que identifique, sintetice y evalúe de forma crítica los estudios que

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

investigaron la fiabilidad del TGMD. En este sentido, solamente se han encontrado estudios de fiabilidad en relación a las dos últimas versiones del TGMD, el TGMD-2 y el TGMD-3.

7.1.1 Fiabilidad test-retest

La fiabilidad test-retest es aquella que refleja la estabilidad de las mediciones realizadas con un instrumento por el mismo evaluador en momentos distintos pero en iguales condiciones[129]. En nuestra revisión sistemática, el primero de los artículos de esta tesis doctoral[144], 10 estudios evaluaron la fiabilidad test-retest del TGMD-2[139,141,143,145–151] y 5 del TGMD-3[139,152–154].En lo que respecta al TGMD-2, se encontraron valores de CCI de buenos a excelentes para la puntuación de las habilidades de control de objetos y la puntuación global del test, y valores de moderados a excelentes en la puntuación de las habilidades locomotrices. El TGMD-3 mostró valores de CCI excelentes para la puntuación global y de buenos a excelentes en las puntuaciones de las habilidades locomotrices y de control de objetos. Basándonos en estos resultados, se pueden considerar ambas versiones del TGMD consistentes en términos de fiabilidad test-retest, ya que la totalidad de los estadísticos recogidos en la revisión sistemática supera los umbrales de fiabilidad aceptable[129,155]. Por otra parte, estos resultados concuerdan con los encontrados en otros instrumentos para la evaluación de la CM orientados al proceso como el Victorian FMS Assessment, e híbridos como el Canadian Agility and Movement Skill Assessment[156].

Por otra parte, se ha mostrado que la familiarización de los participantes evaluados con un instrumento de medida es un factor importante que puede influir en la fiabilidad[157]. En relación con eso, los manuales del TGMD-2[119] y TGMD-3[128] indican que los niños y niñas deben realizar un intento de familiarización tras la descripción verbal y demostración práctica del evaluador. No obstante, la fiabilidad test-retest del TGMD parece ser consistente para ambas variantes a pesar del corto período de familiarización.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7.1.2 Fiabilidad intra-observador

La fiabilidad intra-observador es aquella que muestra el grado de acuerdo entre evaluaciones repetidas realizadas por el mismo evaluador[129]. Nuestra revisión sistemática muestra valores de CCI excelentes para las puntuaciones globales, y valores de buenos a excelentes para la puntuación de las habilidades locomotrices y de control de objetos en ambas variantes del TGMD[144].

Estudios previos reportaron datos de fiabilidad intra-observador de buena a excelente o excelente, utilizando el TGMD-2[141,147,158,159] y TGMD-3[139,152,153,160,161]. En esta línea, centrándonos en los hallazgos del estudio observacional prestando como resultado de esta tesis (*véase Artículo 2 de resultados*)[162], encontramos valores de CCI que indicaron una fiabilidad intra-observador de moderada a excelente, de buena a excelente o excelente para la puntuación global del TGMD y para ambas subescalas. Nuestros resultados de fiabilidad son similares a los observados en los estudios previamente mencionados.

Cabe resaltar que en nuestro estudio se observó que los evaluadores expertos en el TGMD-3, o novatos pero con formación en educación física, obtuvieron resultados más altos de fiabilidad (CCI e intervalos de confianza > 0.75 en todos los casos) al compararlos con los evaluadores sin experiencia ni formación en el ámbito de la educación física (CCI > 0.5)[162]. Esto podría indicar que la formación y experiencia de los evaluadores podría influir en la consistencia de las evaluaciones repetidas en el tiempo por un mismo evaluador utilizando esta batería. Sin embargo, se recomiendan más estudios al respecto para apoyar estas conclusiones debido a la limitada literatura existente.

7.1.3 Fiabilidad inter-observador

La fiabilidad inter-observador nos informa del grado de consistencia de dos o más evaluadores codificando el mismo grupo de sujetos[129]. Los estudios incluidos en la revisión sistemática presentada que evaluaron la fiabilidad inter-observador del TGMD confirmaron niveles adecuados de fiabilidad, oscilando entre de moderada a excelente, de buena a excelente o excelente, tanto para las subescalas locomotriz y de control de objetos como para la puntuación global[144].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Solo uno de los estudios incluidos mostró valores de fiabilidad inter-observador de pobres a moderados[163]. Estos índices de fiabilidad más bajos fueron atribuidos por los autores a la variabilidad observada entre tres habilidades individuales (salto a un pie, salto horizontal y golpeo con raqueta). En línea con nuestros hallazgos, los valores de fiabilidad reportados con otros instrumentos de evaluación de la CM como el Movement Assessment Battery for Children[164], el Bruininks-Oseretsky Test of Motor Proficiency–2nd Edition[137], el Peabody Developmental Motor Scales second edition[137] o el Dragon Challenge[165] fueron similares a los obtenidos en nuestra revisión sistemática.

En el primer estudio observacional presentado en esta tesis (*véase Artículo 2 de resultados*)[162], se encontraron valores de fiabilidad inter-observador de pobre a buena (subescala locomotriz) y de moderada a buena (subescala de control de objetos y puntuación global) usando el TGMD-3. Los índices de fiabilidad encontrados fueron inferiores a los obtenidos por estudios previos[144], lo que puede deberse a la metodología y análisis estadísticos adoptados, así como a los descriptores más exigentes para reportar la fiabilidad empleados en nuestro estudio, como se explicará más adelante. No obstante, los resultados de fiabilidad más bajos no pueden atribuirse en nuestro caso a la experiencia de los evaluadores como ha ocurrido en estudios previos[166], ya que la fiabilidad se mantuvo estable en las comparaciones entre los 5 evaluadores incluyendo los expertos y novatos.

En general, los resultados observados tanto en la revisión sistemática[144] como en el primer estudio observacional de esta tesis[162], muestran valores de fiabilidad inter-observador más bajos que los relativos a la fiabilidad intra-observador, lo que apoya la evidencia de que un evaluador puntúe de forma más consistente consigo mismo que con otros evaluadores[141]. En este sentido, en nuestro segundo estudio observacional (*véase Artículo 3 de resultados*), en el que los evaluadores consensuaron los criterios de rendimiento, se obtuvieron valores de fiabilidad inter-observador al menos de moderados a excelentes, suponiendo esto un incremento de la fiabilidad en comparación el estudio observacional previo[162].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

La variabilidad entre evaluadores al codificar el TGMD ha sido relacionada con la subjetividad y discreción de cada evaluador ya que, aunque existe una descripción para evaluar si se cumplen o no los criterios de rendimiento, la decisión de si el niño o niña evaluado lo cumple estará siempre sujeta al criterio de cada evaluador[167]. Esto provoca que el mismo niño pueda ser evaluado de forma diferente por varios evaluadores[166]. Estos resultados ponen en manifiesto la necesidad de que distintos evaluadores entiendan e interpreten los criterios de rendimiento de igual manera, por ello se ha sugerido que antes de utilizar el TGMD en la práctica clínica los evaluadores lleguen a un acuerdo sobre los criterios de rendimiento para objetivar la evaluación[162].

7.1.4 Fiabilidad por habilidad

Se ha observado que el análisis de fiabilidad de cada habilidad individualmente aporta resultados con mayor variabilidad entre estudios, e incluso contradictorios, siendo además los índices de la fiabilidad de cada habilidad inferiores en la comparativa con las puntuaciones de las subescalas (locomotriz o control de objetos) y puntuación global. Esto podría ser un reflejo de la dificultad de la evaluación de determinados criterios de rendimiento de las HMB[144].

De los veintitrés estudios incluidos en nuestra revisión sistemática, seis[145,153,161,163,168,169] aportaron datos de fiabilidad inter-observador y cuatro[145,153,161,163] de fiabilidad intra-observador de cada habilidad de forma individual. En todos ellos se observaron niveles de fiabilidad intra-observador similares en 2 habilidades locomotrices (skip y salto horizontal) y en 5 de control de objetos (golpeo con raqueta, bote estático, recepción, lanzamiento por encima y lanzamiento por debajo); las seis habilidades restantes mostraron niveles contradictorios de fiabilidad entre los estudios.

En relación con esto, en el análisis realizado por habilidad en los estudios observacionales de la presente tesis doctoral, también se observaron resultados de fiabilidad inferiores y de mayor variabilidad al analizar las habilidades individualmente tanto para la fiabilidad intra- como inter-observador. Esta mayor variabilidad implicaba por un lado, que el mismo evaluador obtuviese resultados de fiabilidad

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

intra-observador pobres en algunas habilidades y excelentes en otras y, por otro, que ante determinadas habilidades algunos evaluadores alcanzasen resultados de fiabilidad pobres y otros buenos o excelentes para la misma habilidad.

En nuestro primer estudio observacional[162], en relación a la fiabilidad intra-observador, se encontraron índices de buenos a excelentes para el galope, bateo y recepción, siendo el desplazamiento lateral, salto con un pie, la patada y el golpeo con raqueta las habilidades con una fiabilidad más débil. Nuestros resultados parecen estar en consonancia con estudios previos[161,163], pero también otras investigaciones obtuvieron resultados contradictorios a los obtenidos en este estudio[145].

En cuanto a la fiabilidad inter-observador, se mostraron en nuestra revisión sistemática[144] índices semejantes en 4 habilidades locomotrices (carrera, galope, zancada y skip) y en 4 habilidades de control de objetos (bote estático, lanzamiento por encima, lanzamiento por debajo rodando y golpeo con raqueta). Las demás habilidades mostraron niveles de fiabilidad contradictorios entre los diferentes estudios. Los resultados del primer estudio observacional de esta tesis[162] muestran valores de pobres a buenos en galope, skip (habilidades locomotrices) y bateo (habilidad de control de objetos), siendo la carrera y salto con un pie las que obtuvieron peores resultados de las habilidades locomotrices y la recepción en las habilidades de control de objetos. Kim et al.[169] y Rintala et al.[163] concuerdan con nuestros resultados en que el galope y el skip son las habilidades con mejores resultados de fiabilidad, y la carrera y el salto con un pie las habilidades con resultados más bajos. Por el contrario, Ayán et al.[145] y Maeng et al.[161] informaron de los resultados de fiabilidad más elevados en el salto con un pie y más débiles en el galope, opuesto a nuestros resultados.

La falta de acuerdo observada entre los diversos estudios hace pensar que la dificultad en la evaluación de las HMB a través de las pruebas orientadas al proceso se encuentra en la manera de entender e interpretar los criterios de rendimiento de cada habilidad individualmente, provocando que cada evaluador cree su propio criterio que no tiene porqué coincidir con el de otros evaluadores[167].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Parece coherente pensar que si los evaluadores llegan a un consenso en cuanto a los criterios de rendimiento antes de las evaluaciones, se reduciría la variabilidad en las puntuaciones y se mejoraría la fiabilidad. Siguiendo este razonamiento, se ha planteado el tercer estudio presentado en esta tesis (*véase Artículo 3 de resultados*) donde se observó que, tras el acuerdo previo a la evaluación de los criterios de rendimiento alcanzado por los evaluadores, la fiabilidad inter-observador se vio incrementada, hecho que se observó tanto a nivel de subescalas como a nivel de habilidades aisladas. Por tanto, podría postularse como alternativa el consenso entre los evaluadores (*véase apartado 7.4 Fiabilidad entre evaluadores*) para reducir el sesgo de subjetividad en la evaluación de la CM con herramientas orientadas al proceso como la TGMD, y para facilitar la comparación de resultados entre investigaciones diferentes.

7.2 EXPERIENCIA DE LOS EVALUADORES

La formación estandarizada para la evaluación utilizando pruebas orientadas al proceso como el TGMD es un requisito para obtener resultados válidos y fiables[161]. En relación con esto, en el manual del TGMD-2[119] y TGMD-3[128] se recomienda la práctica supervisada de al menos tres evaluaciones antes de proceder a la administración e interpretación del TGMD en una situación real. Los resultados de la revisión sistemática presentada en esta tesis muestran que los estudios incluidos no informaron ni describieron de forma sistemática la formación y experiencia de los evaluadores[144]. Por otra parte, se encontró heterogeneidad en la formación académica de los evaluadores en los estudios de fiabilidad incluidos, con perfiles de estudiantes de postgrado en el ámbito de la educación física y ciencias del deporte[145,163,168], fisioterapeutas[150,158], estudiantes de máster[148,170] o estudiantes de doctorado[148,166].

Los hallazgos de nuestro estudio sobre la experiencia de los evaluadores muestran que la fiabilidad inter-observador fue, en la mayoría de las comparaciones, de pobre a buena en las habilidades locomotrices y de control de objetos, así como en las puntuaciones de estas subescalas, y de moderada a buena o excelente para la puntuación global del TGMD-3. Se encontraron limitadas diferencias

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

entre las puntuaciones obtenidas por los evaluadores expertos con más de 5 años de experiencia utilizando el TGMD-3 y los evaluadores novatos (teniendo uno de ellos formación en Educación Física). En contraposición, Palmer et al.[166] mostraron en su estudio que los evaluadores novatos (estudiantes de la Licenciatura de Educación Física con una sesión de formación de dos horas sobre el proceso de codificación) y el experto (estudiante de doctorado en comportamiento motor con más de 3 años de experiencia en la codificación del TGMD-2) puntuaron significativamente distinto las habilidades del TGMD-2 (excepto el galope y la patada).

En cambio, las escasas diferencias significativas encontradas en nuestro estudio tras el análisis de ANOVA de medidas repetidas no parecen ser relevantes. Por un lado, contamos con 5 evaluadores y el total de comparaciones por pares da lugar a 20 análisis en cada subescala y puntuación global del TGMD-3 (10 en la comparación en directo y otros 10 a cámara lenta). De las 60 comparaciones totales, solo se observaron diferencias significativas en 6 de ellas y siempre fue al comparar alguno de los evaluadores novatos con uno de los expertos. Por tanto, la diferencia puede deberse a la forma diferente de entender e interpretar los criterios de rendimiento de cada habilidad por cada evaluador[141,144,146,166,168], más que a la experiencia de los mismos. Por otro lado, las diferencias encontradas entre las puntuaciones de los distintos evaluadores fueron muy pequeñas, un máximo de 6 puntos de diferencia en la puntuación global del TGMD-3, lo que da a entender que no es clínicamente relevante a la hora de evaluar la competencia en las HMB de niños y niñas donde se puede obtener un máximo de 100 puntos.

Estas discrepancias entre nuestro estudio y el de Palmer et al. [166] podrían deberse a que en el citado estudio solo contaron con un evaluador experto y dos novatos y, además, todas las puntuaciones de las HMB de los evaluadores novatos fueron más altas que las del experto. En nuestro caso, hubo 3 evaluadores novatos y 2 expertos y uno de los evaluadores expertos fue el que obtuvo las puntuaciones más altas y, en contraste, uno de los novatos las puntuaciones más bajas. Por último, los participantes del estudio de Palmer et al.[166], estaban en prescolar mientras que nuestros participantes estaban en

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

primaria, lo que podría explicar sus resultados más variables de fiabilidad ya que los niños más pequeños pueden resultar más difíciles de evaluar.

Otros estudios previos que han analizado la fiabilidad inter-observador con instrumentos que requieren protocolos de entrenamiento y preparación similares a los del TGMD, como el Functional Movement Screen[171,172] (para examinar los patrones de movimiento funcionales), el Balance Error Scoring System[173] (para la evaluación de la estabilidad corporal) y el Landing Error Scoring System[174] (para identificar los individuos en riesgo de lesión del ligamento cruzado anterior), encontraron resultados en línea con los nuestros, mostrando puntuaciones de evaluadores novatos aceptables y similares a las de los expertos.

En lo que respecta a la fiabilidad intra-observador, se alcanzó tal y como se mencionó con anterioridad, una fiabilidad al menos de moderada a excelente para las puntuaciones de las subescalas y de la puntuación global del TGMD-3[162]. No obstante, los evaluadores expertos y el evaluador novato con formación en Educación Física consiguieron resultados de fiabilidad más altos que los evaluadores novatos (CCI > 0.9). Esto podría significar que bien la experiencia previa en evaluaciones con el TGMD o la formación más focalizada en la CM podría implicar mayor consistencia en las evaluaciones realizadas por un mismo evaluador.

7.3 EVALUACIÓN EN VÍDEO Y EN DIRECTO

El manual del TGMD no contempla la grabación en vídeo para la evaluación de las habilidades motrices gruesas. No obstante, es el método utilizado habitualmente en los estudios para los análisis de fiabilidad[144]. La utilización de grabaciones en vídeo de niños y niñas realizando las HMB tiene ventajas e inconvenientes. Por un lado, la evaluación a través de vídeo permite que una misma habilidad sea visualizada en repetidas ocasiones para corroborar si se cumplen o no los criterios de rendimiento. Además, permite la observación a cámara lenta de aquellos criterios más complejos de evaluar. Sin embargo, la evaluación mediante vídeos requiere más tiempo que la evaluación en directo, pudiendo no ser adecuada para contextos de evaluación de la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

vida cotidiana en el ámbito clínico o escolar[163]. Por el contrario, la grabación en vídeo podría ser una buena alternativa al permitir la evaluación fuera del horario de consulta o de clase[144].

Otra ventaja de la grabación en vídeo es que los mismos datos pueden ser analizados más de una vez por distintos evaluadores en momentos diferentes, por eso ha sido propuesto como un método adecuado para la investigación[163]. Aunque la grabación en vídeo parece tener más utilidades, no siempre es posible realizarla por consideraciones éticas o por la falta de equipo necesario para llevarla a cabo[168]. Estudios que analizaron la fiabilidad test-retest[143,154] e inter-observador[143,168] del TGMD utilizando la observación en directo mostraron excelentes valores de fiabilidad.

En el estudio presentado como segundo resultado de esta tesis doctoral[162] se analizó la fiabilidad inter- e intra-observador de 5 evaluadores utilizando los dos métodos de evaluación mencionados. Ambos métodos se realizaron evaluando la competencia de las HMB a través de vídeos (para que fuese posible la evaluación por todos los evaluadores que residían en diferentes zonas del territorio nacional e incluso internacional). La evaluación "en directo" fue simulada, visualizando el vídeo a velocidad normal, en una sola ocasión y sin pausa; este método ya fue utilizado con el mismo fin en estudios previos para evaluar la calidad del movimiento[172]. Por otro lado, en la observación "en vídeo" la grabación podía ser vista tantas veces como fuese necesaria, a cámara lenta y con pausa. Se obtuvo una fiabilidad intra-observador al menos de moderada a excelente con ambos métodos de visualización tanto en las puntuaciones de las subescalas como en la puntuación global del TGMD. Aunque el análisis post-hoc de Bonferroni reveló diferencias significativas entre las puntuaciones obtenidas entre las observaciones en directo y a cámara lenta de algún evaluador (sobre todo de uno de los evaluadores novatos), fueron mínimas, suponiendo una diferencia de 1 a 3 puntos (de un total de 100 puntos del test) entre las puntuaciones establecidas en evaluación en directo *vs* en vídeo.

En este sentido, en el estudio de Shultz et al[175] en el que analizaron la fiabilidad intra-observador del Functional Movement Screen (test para observar los patrones de movimiento funcionales)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

con los dos métodos de observación concluyeron que ambos métodos son comparables al presentar una fiabilidad excelente (CCI > 0.9).

De igual forma, la fiabilidad inter-observador no se vio afectada en nuestro estudio al comparar los dos modos de visualización. De acuerdo con los resultados expuestos, la fiabilidad del TGMD no parece estar influenciada por los modos de observación, obteniéndose resultados de fiabilidad aceptables con ambos métodos[144,162]. No obstante, según nuestro conocimiento, este es el primer estudio en comparar la fiabilidad en directo y en video, por tanto, sería recomendable seguir investigando en esta línea.

7.4 CONSENSO ENTRE EVALUADORES

Se ha destacado la necesidad de un mayor entrenamiento y una aclaración de los criterios de rendimiento, tanto de las habilidades locomotrices[166] como de control de objetos[168], entre los evaluadores antes de utilizar el TGMD en una situación real. En relación con esto, se ha propuesto el tercer estudio presentado en esta tesis partiendo de la premisa de la importancia de reducir el sesgo de subjetividad de los criterios de rendimiento de las habilidades entre los evaluadores. Se observó que, tras un proceso de consenso entre tres evaluadores sobre los criterios de rendimiento de las habilidades del TGMD-3, la fiabilidad inter-observador mejoró tanto en las habilidades locomotrices como en las de control de objetos.

Estudios previos[141,167] destacaban la posibilidad de que la subjetividad de los criterios de rendimiento provoque que cada evaluador los entienda e interprete de forma distinta, siendo más probable que esté de acuerdo de forma consistente consigo mismo que con otros evaluadores, razón por la que la fiabilidad intra-observador suele ser superior a la inter-observador[144,162]. En línea con esto, en el proceso de consenso entre los evaluadores de nuestro estudio se demostró esta hipótesis formulada por otros autores y autoras ya que, ante el mismo criterio de rendimiento lo puntuaban de forma distinta. Por ejemplo, en la habilidad de salto con un pie, en el segundo criterio "*el pie de la pierna que no salta (pierna libre) permanece detrás del cuerpo*" algún evaluador lo interpretaba de forma literal (es decir, que tenía que permanecer durante todo el tiempo de ejecución de la

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

habilidad el pie detrás del cuerpo) y otros eran más permisivos (es, decir, si la mayoría del tiempo lo mantenían detrás aunque en algún momento sobrepasase el cuerpo, lo puntuarían favorablemente), hecho que hace imprescindible un proceso de consenso previo a la evaluación.

En relación con las habilidades locomotrices, antes del consenso, los evaluadores de nuestro estudio alcanzaron una fiabilidad aceptable (CCI y límite inferior del intervalo de confianza ≥ 0.5) en 2 de 6 habilidades y tras el consenso, en 4 de las 6 habilidades, suponiendo una mejora del 50%. Las habilidades que permanecieron con una baja fiabilidad fueron la carrera y el salto con un pie. Tanto la carrera[153,161,162,169] como el salto con un pie[153,162,163] son habilidades que han mostrado mayor complejidad en investigaciones previas. Parece que estas habilidades tienen una subjetividad y complejidad adicional que provoca valores de fiabilidad más bajos entre los evaluadores a pesar del consenso. En cuanto a la carrera puede deberse a que, al realizar la evaluación mediante vídeos, la cámara tuvo que colocarse a una distancia suficiente como para grabar 20m de carrera y por tanto en una distancia tan amplia resulta complicado la codificación de criterios como "*el contacto del pie se reduce al talón o punta (no con el pie plano)*". Esto implica que más allá del consenso, este criterio pueda codificarse de forma diferente por la falta de precisión en la evaluación. En lo que respecta al salto con un pie, dos de sus criterios "*La pierna que no salta se balancea hacia delante de forma pendular para producir fuerza*" y "*Los brazos se flexionan y se balancean hacia delante para producir fuerza*" utilizan la descripción "…*para producir fuerza*" para determinar si la ejecución se realiza correctamente. Esto requiere, por tanto, que los evaluadores diferencien entre el balanceo como un movimiento para el mantenimiento del equilibrio y el balanceo específicamente para producir fuerza.

Por otro lado, los evaluadores de nuestro estudio pasaron de una fiabilidad aceptable (CCI y límite inferior del intervalo de confianza ≥ 0.5) en 1 de las 7 habilidades de control de objetos antes del consenso, a 4 de las 7 habilidades tras el mismo. De nuevo, la mejora fue notable, aunque el bote estático, la patada y la recepción se

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mantuvieron con una baja fiabilidad. Estos resultados no son sorprendentes y son similares a los aportados por investigaciones previas que encontraron el bote estático[162], la patada[161–163,169] y la recepción[162,168] como habilidades con peor fiabilidad. De nuevo, esto podría deberse a la mayor dificultad de los criterios de rendimiento que necesitarán más atención en el futuro. Por ejemplo, en el caso de la patada uno de los criterios indica "*Pie de apoyo colocado cerca del balón*" sin embargo "*cerca*" puede ser interpretado como pegado al balón o a cierta distancia.

En el caso de la puntuación de las subescalas y la puntuación global, las mejoras fueron menos notables porque la fiabilidad suele ser menos variable que al analizar las habilidades de forma individual, tal y como se explicó en el apartado *7.1.3 Fiabilidad por habilidad*. No obstante, tras el consenso de los evaluadores, la fiabilidad aumentó especialmente en las puntuaciones de la subescala locomotriz, que mejoró de una fiabilidad inter-observador de pobre a buena a una de moderada a excelente.

En base a estos resultados, se recomienda que los evaluadores establezcan un acuerdo previo a la evaluación para objetivar los distintos puntos de vista, interpretaciones y métodos de evaluación y conseguir una codificación más fiable. Además, las futuras investigaciones deberán prestar mayor atención en las habilidades de carrera, salto con un pie, bote estático, recepción y patada por ser aquellas con resultados más inconsistentes a pesar del acuerdo. Por último, sería recomendable llegar a un consenso común y que éste fuese ampliamente difundido entre todos los usuarios del TGMD, de forma que la fiabilidad no solo mejore entre los miembros de un grupo de investigación, sino también entre distintos equipos de investigación y profesionales clínicos.

7.5 ANÁLISIS ESTADÍSTICO DE FIABILIDAD

Existe una amplia variabilidad de procedimientos estadísticos empleados para determinar la fiabilidad[144]. El hecho de que exista esta gran heterogeneidad a la hora de seleccionar el procedimiento estadístico (entre los más usados se encuentran el Coeficiente de Correlación Intraclase, Coeficiente de Correlación de Pearson, Kappa

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

o el Índice de Validez de Contenido) dificulta la síntesis y comparación entre los estudios[144].

En general, los resultados de fiabilidad del TGMD obtenidos en los estudios realizados en esta tesis doctoral son inferiores a los reportados en la revisión sistemática presentada[144], lo que puede tener relación con la metodología adoptada o el análisis estadístico utilizado. Por un lado, se ha recomendado la utilización del Coeficiente de Correlación Intraclase (CCI) como estadístico ideal para el análisis de la fiabilidad por reflejar el grado de correlación y acuerdo entre las mediciones[129,161]; no obstante, numerosos estudios siguen utilizando el Coeficiente de Correlación de Pearson[143,145,148,150,152,153,158], Kappa[151] o el Índice de Validez de Contenido[167].

Por otra parte, cuando se emplea el CCI hay que seleccionar un modelo (modelo de efectos aleatorios unidireccional, modelo de efectos aleatorios bidireccional o modelo de efectos mixtos bidireccional) la definición de CCI (acuerdo absoluto o consistencia) y el tipo de CCI (medidas únicas o medias promedio) según el estudio concreto. Sin embargo, algunos estudios no informan del modelo/tipo/definición del CCI, siendo esto inapropiado. Además, como no hay valores estándares para informar de la fiabilidad, numerosos estudios utilizan descriptores menos exigentes que los empleados en esta tesis, la cual siguió las recomendaciones de Koo y Li[129]. Según esto, en nuestros estudios se consideraron los valores superiores a 0.75 como buenos y los superiores a 0.90 como excelentes. Otros estudios consideraron los valores de CCI superiores a 0.6 como buenos y los superiores a 0.75 como excelentes[139,141,152,159–161,163,169]. De haber utilizado estos valores para interpretar el CCI, hubiésemos obtenido valores más altos de fiabilidad.

Por último, la fiabilidad debe interpretarse además de con el estadístico CCI con el límite superior e inferior del intervalo de confianza del 95%[129], ya que los valores del CCI no describen la fiabilidad en sí mismos, otra razón que podría explicar nuestros resultados de fiabilidad más bajos que los de otros estudios que utilizaron solo el valor del CCI[141,145,148,153,158,160,161,163,168,169].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sería recomendable de cara a futuras investigaciones que se llegase a una estandarización tanto de los estadísticos empleados como de las interpretaciones de cada uno para facilitar la comparación entre diferentes estudios sobre la evaluación de la fiabilidad del TGMD.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8 LIMITACIONES Y FORTALEZAS

8.1 LIMITACIONES

Los estudios incluidos en la tesis doctoral no están libres de limitaciones, que deben ser consideradas. En primer lugar, en lo que respecta a la revisión sistemática, en los criterios de elegibilidad se estableció la exclusión de la llamada *literatura gris* (es decir, monografías, resúmenes de conferencias y tesis) y, además, sólo se seleccionaron las publicaciones en inglés, español o portugués que investigaban como objetivo principal la fiabilidad del TGMD. Por tanto, publicaciones relevantes pudieron no ser incluidas en la revisión.

Por otro lado, en los dos estudios observacionales presentados, el análisis de fiabilidad se realizó evaluando a 25 niños y niñas aparentemente sanos, lo que podría considerarse una muestra reducida y, por tanto, los resultados deben interpretarse con precaución. Se recomienda que nuestros resultados sean replicados por futuros estudios con más participantes y de perfil más diverso (p. ej.: diferentes orígenes socioeconómicos, personas con discapacidad).

En lo que respecta al tercer estudio concretamente, para analizar la fiabilidad inter-observador tras el acuerdo de los evaluadores, la ejecución de las HMB fue codificada por tercera vez por los evaluadores (anteriormente para el análisis la fiabilidad inter- e intra-observador antes del consenso). Sin embargo, como se ha establecido que los evaluadores pueden mantener el sesgo de memoria después de un período de lavado de 14 días[128], introdujimos un intervalo de 9 meses en el presente estudio para evitar este posible factor de confusión.

Por otra parte, la amplia variedad de procedimientos estadísticos empleados para determinar la fiabilidad y la heterogeneidad de los participantes dificulta la comparación de resultados con otros estudios

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

y, en el caso de la revisión sistemática, impidió la realización de un meta-análisis. Por tanto, sería recomendable y valioso estandarizar este enfoque.

8.2 FORTALEZAS

La revisión sistemática presentada ha sido la primera en identificar y evaluar la evidencia científica sobre la fiabilidad del TGMD, sintetizando los resultados de consistencia interna, test-retest y fiabilidad inter- e intra-observador en un mismo documento.

Los dos estudios observacionales, además de haber utilizado un instrumento psicométrico robusto para evaluar las HMB, aportan resultados novedosos sobre la fiabilidad de los evaluadores utilizando el TGMD-3. Por un lado, se ha estudiado la fiabilidad inter e intra-observador utilizando 5 evaluadores (3 de ellos novatos y 2 expertos). Solo un estudio previo analizó la fiabilidad con 5 evaluadores[161] y ninguno comparando la experiencia de los mismos. Además, fue el primer estudio que analizó la influencia de la evaluación en vídeo y en directo en la fiabilidad. Por otro lado, se aportan datos innovadores de fiabilidad tras el acuerdo alcanzado por los evaluadores sobre los criterios de rendimiento de las HMB del TGMD-3. Estos resultados podrían servir como base para futuras investigaciones en las que se quiera realizar una evaluación más fiable de las HMB en niños y niñas o en el ámbito investigador para continuar la línea de los estudios de fiabilidad.

Por último, que el desarrollo de *Alfamov* se haya realizado con un equipo multidisciplinar compuesto por profesionales sanitarios, de la educación física, de diseño gráfico y audiovisual e informáticos aporta un valor añadido en la creación del instrumento por los distintos y valiosos puntos de vista proporcionados por los profesionales. Por tanto, *Alfamov* se propone como un instrumento adecuado para la evaluación de las HMB en diferentes contextos.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

9 IMPLICACIONES PRÁCTICAS

La revisión sistemática y los estudios observacionales realizados han puesto de manifiesto una serie de cuestiones que deben ser consideradas por la comunidad científica y por las personas que quieran evaluar las HMB utilizando el TGMD en un entorno práctico:

- El TGMD en sus dos versiones (TGMD-2 y TGMD-3) ha mostrado ser un instrumento robusto, adecuado para la evaluación de las HMB, por tanto puede ser utilizado con confianza para esta tarea en la práctica.
- La fiabilidad inter-observador no se ve influenciada por la experiencia de los evaluadores (novatos y expertos), poniendo de manifiesto la posibilidad de su uso en evaluadores de distintos perfiles, lo que facilitará la evaluación y estandarización de las evaluaciones.
- La fiabilidad intra-observador fue más elevada en los evaluadores expertos o novatos con formación en Educación Física, significando esto que los evaluadores novatos sin formación específica en el ámbito de la CM pueden necesitar de una familiarización mayor del instrumento para ser más consistentes consigo mismos.
- La fiabilidad inter- e intra-observador no varía dependiendo del tipo de observación que se utilice (en vídeo o en directo), siendo ambos métodos igual de fiables, lo que debe ser tenido en cuenta para decidir qué tipo de observación se debe/puede aplicar en cada situación concreta.
- Se recomienda la elaboración de un consenso sobre los criterios de rendimiento de las HMB ya que esto produce un aumento de fiabilidad inter-observador, reduciendo el riesgo de subjetividad de interpretación de cada evaluador.

Por otro lado, distintas investigaciones han destacado la necesidad de creación de herramientas digitales para la evaluación de

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

las HMB[76,81,105] que incluyan habilidades locomotrices, manipulativas o control de objetos y de estabilidad[105], y que puedan ser interpretadas y utilizadas por personas con menos conocimiento específico de este campo concreto (por ejemplo el profesorado de Educación Primaria sin especialidad en Educación Física)[76].

De acuerdo con esto, la herramienta digital presentada en los resultados, *Alfamov*, podría ser una buena alternativa para la evaluación de la CM en entornos escolares y clínicos. Al contener una parte didáctica (que cuenta con vídeos y explicaciones detalladas sobre cada habilidad), podría suplir la falta de conocimiento que han manifestado previamente los docentes como una de las razones principales por las que no evalúan las HMB[109]. Además, *Alfamov* incluye 22 HMB de las tres subescalas (locomotrices, manipulativas o control de objetos y de estabilidad), lo que suple la limitación que supone la ausencia de las habilidades de estabilidad en otras baterías[104]. En relación con esto, estas 22 habilidades surgen de la recopilación de 4 test[80,117,119,128,134] de evaluación de HMB ya que no hay suficiente evidencia para justificar el uso de una batería concreta[105]. *Alfamov* es también una herramienta útil para la investigación que puede ser usada por investigadores principalmente del ámbito educativo y clínico que trabajen sobre la CM.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10 CONCLUSIONES

Las conclusiones han sido determinadas en base a las hipótesis formuladas:

H.1 – Existe suficiente literatura científica sobre la fiabilidad del *Test of Gross Motor Development (TGMD)* para realizar una valoración crítica, un resumen de los resultados y una recomendación del test como instrumento robusto para la evaluación de las habilidades motrices básicas.

C.1 – **La revisión sistemática realizada sobre la fiabilidad del TGMD, mostró suficiente evidencia para poder considerar la mencionada prueba orientada al proceso como un instrumento psicométrico robusto y apropiado para la evaluación de las HMB.**

En general, los resultados indican que, independientemente de la variante de la prueba (TGMD-2 o TGMD-3) y del tipo de evaluación (en directo o en vídeo), el TGMD tiene una consistencia interna de buena a excelente para la puntuación global y de moderada a excelente para las subescalas locomotriz y de control de objetos. También cuenta con una fiabilidad inter-observador de buena a excelente (para la puntuación de las subescalas locomotriz, de control de objetos y puntuación global). La fiabilidad intra-observador fue de buena a excelente para ambas subescalas y excelente para la puntuación global. En cuanto al test-retest se alcanzó una fiabilidad de buena a excelente en la puntuación de la subescala de control de objetos y en la puntuación global y de moderada a excelente en la subescala locomotriz. Además, la fiabilidad del TGMD parece ser alta tanto en niños con desarrollo típico como en niños con discapacidades; sin embargo, el menor número de estudios en poblaciones especiales revela la necesidad de continuar investigando en esta línea.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

H.2 – La fiabilidad inter- e intra-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* entre evaluadores expertos es superior a la de los evaluadores novatos y mediante la evaluación mediante vídeo es superior a la evaluación en directo.

C.2 – **El análisis de fiabilidad mostró una mayor fiabilidad intra-observador, para las puntuaciones de las habilidades locomotrices, de control de objetos y puntuaciones globales, en los evaluadores expertos y el evaluador novato con formación en Educación Física (fiabilidad de buena a excelente *vs* de moderada a excelente en los evaluadores novatos). No obstante, la fiabilidad inter-observador no pareció estar influenciada por la formación y la experiencia de los evaluadores manteniéndose los resultados de fiabilidad estables entre evaluadores expertos y novatos.** Los modos de evaluación (en directo o en vídeo) tampoco parecieron influir en la fiabilidad, pero es necesario seguir investigando en esta línea ya que fue el primer estudio en aportar estos resultados.

Se encontró una mayor variabilidad en la fiabilidad intra- e inter-observador cuando se analizaron las habilidades por separado. Además, los valores más altos de fiabilidad intra-observador que de inter-observador reflejan que los criterios de rendimiento de las HMB son interpretados de forma diferente por cada evaluador. Por ello, para la práctica clínica sería recomendable que los evaluadores llegaran a un acuerdo antes de la evaluación para evitar interpretaciones subjetivas que pudieran distorsionar los resultados, e incluso revisar los criterios de rendimiento para objetivarlos más.

H.3 – La fiabilidad inter-observador del *Test of Gross Motor Development – Third Edition (TGMD-3)* mejora tras un acuerdo previo entre los evaluadores de los criterios de rendimiento de cada habilidad.

C.3 – **Tras el acuerdo de los evaluadores sobre los criterios de rendimiento, mejoró en general la fiabilidad inter-observador de las habilidades del TGMD-3. Esta mejora se observó sobre todo**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

en las habilidades locomotrices en las que se pasó de una fiabilidad de pobre a buena antes del consenso, a una fiabilidad de moderada a excelente tras el mismo. Por tanto, se recomienda que tanto en el ámbito de la investigación como en el ámbito clínico cuando se use el TGMD-3 como herramienta de evaluación de las HMB en niños y niñas, se debe prestar atención a la elaboración de un acuerdo previo sobre los criterios de rendimiento cuando estén inmersos varios evaluadores con el fin de reducir el sesgo de subjetividad entre ellos. Se sugiere que se preste especial atención a las habilidades más complejas, es decir, la carrera, salto con un pie, bote estático, recepción y patada. Además, sería recomendable que se compartan y difundan los criterios de rendimiento mejorados en cada estudio para permitir un consenso común y una formación específica y estandarizada para todos los usuarios del TGMD.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

11 CONCLUSIONS

The conclusions have been established based on the hypotheses formulated:

H.1 - There is enough scientific literature on the reliability of the Test of Gross Motor Development (TGMD) to do a critical review, a summary of the results, and a recommendation of the test as a robust tool for fundamental movement skills (FMS) assessment.

C.1 - **The systematic review of the reliability of the TGMD showed sufficient evidence to consider the TGMD as a robust and appropriate psychometric instrument for assessment of FMS**.

Overall, the results indicate that regardless of the test variant (TGMD-2 or TGMD-3) and the type of assessment (live or video), the TGMD has good-to-excellent internal consistency for the global score and moderate-to-excellent internal consistency for the locomotor and ball skills subscales. It also has good-to-excellent inter-rater reliability (for the locomotor and ball skills subscales and global score). Intra-rater reliability was good-to-excellent for both subscales and excellent for the global score. Test-retest reliability was good-to-excellent for the ball skill subscale and the global score and moderate-to-excellent for the locomotor subscale. In addition, the reliability of the TGMD appears to be high in both typically developing children and children with disabilities; however, the smaller number of studies in special populations reveals the need for further research in this area.

H.2 - The inter- and intra-rater reliability of the Test of Gross Motor Development - Third Edition (TGMD-3) between expert raters is higher than that of novice raters, and the reliability using video assessment is higher than using the live one.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

C.2 - **Reliability analysis showed higher intra-rater reliability, for locomotor skills scores, ball skills scores and global scores, in the expert raters and the novice rater with PE training (good-to-excellent vs moderate-to-excellent reliability in the novice raters). However, inter-rater reliability did not seem to be influenced by the training and experience of the raters, with reliability scores remaining stable between expert and novice raters.** The assessment methods (live or video) did not seem to influence reliability either, but further research is needed in this area as this is the first study to provide such results.

Higher variability in intra- and inter-rater reliability was found when skills were analysed separately. Furthermore, the higher values for intra-rater than inter-rater reliability reflect the fact that the FMS performance criteria are interpreted differently by every rater. Therefore, for clinical practice, it would be recommended for the raters to reach an agreement before the assessment in order to avoid subjective interpretations that could distort the results; and even to revise the performance criteria to make them more objective.

H.3 - The inter-rater reliability of the Test of Gross Motor Development - Third Edition (TGMD-3) improves after the prior agreement between raters on the performance criteria for each skill.

C.3 - **After rater agreement on the performance criteria, the inter-rater reliability of the TGMD-3 skills improved overall. This improvement was observed mainly in locomotor skills, where reliability went from poor-to-good before agreement to moderate-to-excellent after it.** Therefore, it is recommended that in both research and clinical settings, when using the TGMD-3 as an assessment tool for FMS in children, attention should be paid to developing prior agreement on performance criteria when several raters are involved in order to reduce subjectivity bias among them. It is suggested that special attention be paid to the more complex skills, i.e. running, hopping, dribbling, catching and kicking. In addition, it would be advisable to share and disseminate the improved performance criteria in each study to allow for a common consensus and specific and standardised training for all TGMD users.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

12 LÍNEAS FUTURAS DE INVESTIGACIÓN

Se ha destacado a lo largo de este trabajo la necesidad de más estudios de fiabilidad del TGMD en poblaciones especiales (personas con TEA, con trastorno por déficit de atención e hiperactividad, con deficiencias visuales, con trastornos mentales y conductuales o con discapacidad intelectual). Por otro lado, se requieren más estudios de fiabilidad con evaluadores de diferentes perfiles y experiencia ya que en la práctica real es posible que diferentes personas (profesorado de preescolar, primaria, personal de pediatría, enfermería) y sin experiencia previa necesiten evaluar la competencia de las HMB. También, como se ha destacado, esta investigación ha sido la primera en analizar las posibles diferencias entre la evaluación de las HMB por medio del TGMD-3 en vídeo y en directo. Por tanto, y de acuerdo a las limitaciones ya mencionadas de nuestro diseño, sería recomendable realizar más estudios que confirmen los hallazgos encontrados.

En lo que respecta a la necesidad de operativizar la evaluación de las HMB desde la perspectiva de reducir el tiempo y trabajo de evaluación, recientemente se han estudiado versiones reducidas del TGMD en niños y niñas de Brasil[136,176,177] e Irlanda[178]. No obstante, estos estudios difieren en las habilidades seleccionadas para la versión reducida, destacando que posiblemente esto sea debido a las diferencias culturales existentes. Por ello, enfatizada la necesidad de la evaluación más eficiente de la CM[177,178], sobre todo en lo que se refiere al tiempo empleado[84,136], futuros estudios deberían investigar versiones reducidas de herramientas para la evaluación de las HMB, preferiblemente que incluyan las tres subescalas[80,105] (locomotrices, manipulativas y de estabilidad) en distintas poblaciones.

En ese sentido, por un lado, sería adecuado y conveniente probar *Alfamov* en distintos entornos y por diferentes profesionales para analizar posibles mejoras que necesiten ser implementadas para

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

operativizar la evaluación de las HMB. Tras esto, sería recomendable estudiar la posibilidad de una versión reducida de *Alfamov* que permitiese recomendar a investigadores, profesorado y profesionales de la salud las habilidades concretas que deberían evaluar en base a aquellas que mejor expliquen la puntuación total de CM teniendo en cuenta el sexo y edad de los niños y niñas. Por otra parte, *Alfamov* podría ser implementada en el sistema educativo y sanitario posibilitando la conexión entre ambos para trabajar de forma coordinada en el desarrollo y control de la CM de los niños y niñas.

Por último, como se ha mencionado, las diferencias culturales existentes entre las distintas regiones provocan diferencias en la CM de niños y niñas[50]. Por ejemplo, es posible que los estándares normativos de desarrollo de las HMB de Estados Unidos no reflejen la realidad de la CM de niños y niñas europeos[104]. No obstante, no existen estándares normativos adaptados a nuestra población. Por esta razón, sería necesario y recomendable la creación de estándares normativos de competencia de las HMB adaptados a la población española.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

13 REFERENCIAS BIBLIOGRÁFICAS

1. World Health Organization. Physical Activity [Internet]. 2022 [citado 2 de octubre de 2022]. Disponible en: https://www.who.int/news-room/fact-sheets/detail/physical-activity.

2. World Health Organization. Global Action Plan on Physical Activity 2018-2030. More Active People for a Healthier World. Geneva, Switzerland; 2018.

3. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act. 2017;14(1):1-17.

4. Sedentary Behaviour Research Network. Standardized use of the terms «sedentary» and «sedentary behaviours». Appl Physiol Nutr Metab. 2012;37(3):540-2.

5. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1,9 million participants. Lancet Glob Heal. 2018;6(10):e1077-86.

6. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1,6 million participants. Lancet Child Adolesc Heal. 2020;4(1):23-35.

7. Pan American Health Organization. World Health Organization. Noncommunicable Diseases [Internet]. [citado 29 de septiembre de 2022]. Disponible en: https://www.paho.org/en/topics/noncommunicable-diseases.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Impact of Physical Inactivity on the World's Major Non-Communicable Diseases. Child Dev. 2012;380(9838):219-29.

9. Schmid D, Leitzmann MF. Television viewing and time spent sedentary in relation to cancer risk: A meta-analysis. J Natl Cancer Inst. 2014;106(7):1-19.

10. Sattelmair J, Pertman J, Ding E, Kohl H, Haskell W, Lee I. Dose-Response Between Physical Activity and Risk of Coronary Heart Disease: A Meta-Analysis. Circulation. 2011;124(7):789-95.

11. Guo C, Zhou Q, Zhang D, Qin P, Li Q, Tian G, et al. Association of total sedentary behaviour and television viewing with risk of overweight/obesity, type 2 diabetes and hypertension: A dose–response meta-analysis. Diabetes, Obes Metab. 2020;22(1):79-90.

12. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: Systematic review and meta-analysis. Diabetologia. 2012;55(11):2895-905.

13. Löllgen H, Böckenhoff A, Knapp G. Physical activity and all-cause mortality: An updated meta-analysis with different intensity categories. Int J Sports Med. 2009;30(3):213-24.

14. Hrafnkelsdottir SM, Brychta RJ, Rognvaldsdottir V, Gestsdottir S, Chen KY, Johannsson E, et al. Less screen time and more frequent vigorous physical activity is associated with lower risk of reporting negative mental health symptoms among Icelandic adolescents. PLoS One. 2018;13(4):1-15.

15. Pengpid S, Peltzer K. Sedentary behaviour, physical activity and life satisfaction, happiness and perceived health status in university students from 24 countries. Int J Environ Res Public Health. 2019;16(12): 2084.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

16. World Health Organization. Obesity and overweight [Internet]. 2021 [citado 8 de septiembre de 2022]. Disponible en: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

17. WHO Consultation on Obesity, World Health Organization. Obesity: preventing and managing the global epidemic: report of a WHO consultation. [Internet]. Geneva, Switzerland; 2000. Disponible en: https://apps.who.int/iris/handle/10665/42330.

18. Estudio ALADINO 2019: Estudio sobre Alimentación, Actividad Física, Desarrollo Infantil y Obesidad en España 2019. Madrid: Agencia Española de Seguridad Alimentaria y Nutrición. Ministerio de Consumo; 2020.

19. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451-62.

20. Varkevisser RDM, van Stralen MM, Kroeze W, Ket JCF, Steenhuis IHM. Determinants of weight loss maintenance: a systematic review. Obes Rev. 2019;20(2):171-211.

21. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland; 2010.

22. Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. J Psychiatr Res. 2016;77:42-51.

23. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet. 2017;390(10113):2673-734.

24. Das P, Horton R. Rethinking our approach to physical activity. Lancet. 2012;380(9838):189-90.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

25. Warburton DER, Bredin SSD. Health benefits of physical activity: A systematic review of current systematic reviews. Curr Opin Cardiol. 2017;32(5):541-56.

26. Macarro Moreno J, Romero Cerezo C, Torres Guerrero J. Motivos de abandono de la práctica de actividad físico-deportiva en los estudiantes de bachillerato de la provincia de Granada. Rev Educ. 2010;353:495-519.

27. Crane J, Temple V. A systematic review of dropout from organized sport among children and youth. Eur Phys Educ Rev. 2015;21(1):114-31.

28. Henderson S, Sugden D. Movement assessment battery for children. London: The Psychological Corporation; 1992.

29. Witt PA. Why Children/Youth Drop Out of Sports. J Park Recreat Admi. 2018;36(3):191-9.

30. Lopes VP, Rodrigues LP, Maia JAR, Malina RM. Motor coordination as predictor of physical activity in childhood. Scand J Med Sci Sport. 2011;21(5):663-9.

31. Cairney J, Kwan MYW, Velduizen S, Hay J, Bray SR, Faught BE. Gender, perceived competence and the enjoyment of physical education in children: A longitudinal examination. Int J Behav Nutr Phys Act. 2012;9:1-8.

32. Da Silva DRP, Werneck AO, Collings P, Fernandes RA, Ronque ERV, Sardinha LB, et al. Identifying children who are susceptible to dropping out from physical activity and sport: A cross-sectional study. Sao Paulo Med J. 2019;137(4):329-35.

33. Vella SA, Cliff DP, Okely AD. Socio-ecological predictors of participation and dropout in organised sports during childhood. Int J Behav Nutr Phys Act. 2014;11(1):1-10.

34. King-Dowling S, Proudfoot NA, Cairney J, Timmons BW. Motor Competence, Physical Activity, and Fitness across Early Childhood. Med Sci Sports Exerc. 2020;52(11):2342-8.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

35. Lubans DR, Morgan PJ, Cliff DP, Barnett LM, Okely AD. Fundamental movement skills in children and adolescents: Review of associated health benefits. Sport Med. 2010;40(12):1019-35.

36. Barnett LM, Morgan PJ, van Beurden E, Beard JR. Perceived sports competence mediates the relationship between childhood motor skill proficiency and adolescent physical activity and fitness: A longitudinal assessment. Int J Behav Nutr Phys Act. 2008;5:1-12.

37. Alert MD, Saab PG, Llabre MM, McCalla JR. Are Self-Efficacy and Weight Perception Associated With Physical Activity and Sedentary Behavior in Hispanic Adolescents? Heal Educ Behav. 2019;46(1):53-62.

38. Liu M, Wu L, Ming Q. How does physical activity intervention improve self-esteem and self-concept in children and adolescents? Evidence from a meta-analysis. PLoS One. 2015;10(8):1-17.

39. International Physical Literacy Association. [Internet]. 2017 [citado 8 de septiembre de 2022]. Disponible en: https://www.physical-literacy.org.uk.

40. Whitehead M. Definitions of Physical Literacy: developments and Issues. En: Whitehead M, editor. Physical Literacy across the World. London, UK: Routledge; 2019. p. 8-18.

41. Jurbala P. What Is Physical Literacy, Really? Quest. 2015;67(4):367-83.

42. Subdirección General de Cooperación Territorial e Innovación Educativa Gobierno de España. Alfabetizaciones Múltiples [Internet]. [citado 9 de septiembre de 2022]. Disponible en: https://www.educacionyfp.gob.es/mc/sgctie/alfabetizaciones-multiples.html.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

43. Whitehead M. Physical literacy. En: Congress of the International Association of Physical Education and Sport for Girls and Women. Melbourne, Australia; 1993.

44. Whitehead M. Physical Literacy: Throughout the lifecourse. Whitehead M, (Ed.). London, UK: Routledge; 2010.

45. Edwards LC, Bryant AS, Keegan RJ, Morgan K, Jones AM. Definitions, Foundations and Associations of Physical Literacy: A Systematic Review. Sport Med. 2017;47(1):113-26.

46. Dudley D, Cairney J, Wainwright N, Kriellaars D, Mitchell D. Critical Considerations for Physical Literacy Policy in Public Health, Recreation, Sport, and Education Agencies. Quest. 2017;69(4):436-52.

47. Shearer C, Goss HR, Edwards LC, Keegan RJ, Knowles ZR, Boddy LM, et al. How is physical literacy defined? A contemporary update. J Teach Phys Educ. 2018;37(3):237-45.

48. Stodden DF, Goodway JD, Langendorfer SJ, Roberton MA, Rudisill ME, Garcia C, et al. A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity: An Emergent Relationship. Quest. 2008;60(2):290-306.

49. Logan SW, Kipling Webster E, Getchell N, Pfeiffer KA, Robinson LE. Relationship Between Fundamental Motor Skill Competence and Physical Activity During Childhood and Adolescence: A Systematic Review. Kinesiol Rev. 2015;4(4):416-26.

50. Hulteen RM, Morgan PJ, Barnett LM, Stodden DF, Lubans DR. Development of Foundational Movement Skills: A Conceptual Model for Physical Activity Across the Lifespan. Sport Med. 2018;48(7):1533-40.

51. Cairney J, Dudley D, Kwan M, Bulten R, Kriellaars D. Physical Literacy, Physical Activity and Health: Toward an Evidence-Informed Conceptual Model. Sports Med. 2019;49(3):371-83.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

52. Barnett LM, Webster EK, Hulteen RM, De Meester A, Valentini NC, Lenoir M, et al. Through the Looking Glass: A Systematic Review of Longitudinal Evidence, Providing New Insight for Motor Competence and Health. Sports Med. 2022;52(4):875-920.

53. Sport Australia. Australian Physical Literacy Framework. Australian Sports Commission, (Ed.). 2019.

54. Logan SW, Ross SM, Chee K, Stodden DF, Robinson LE. Fundamental motor skills: A systematic review of terminology. J Sports Sci. 2018;36(7):781-96.

55. Sport for Life. Physical Literacy [Internet]. 2019 [citado 9 de septiembre de 2022]. Disponible en: https://sportforlife.ca/10-key-factors/physical-literacy/.

56. Higgs C, Balyi I, Way R, Cardinal C, Norris S, Bluechardt M. Developing Physical Literacy A Guide For Parents Of Children Ages 0 to 12 [Internet]. Canadian Sport for Life Physical Literacy. Vancouver, Canada: Canadian Sport Centres; 2008. Disponible en: https://sportforlife.ca/portfolio-view/developing-physical-literacy-a-guide-for-parents-of-children-ages-0-to-12/.

57. Richards R. Physical Literacy and Sport. Canberra: Australian Sports Commission; 2016.

58. Physical and Health Education Canada. About us [Internet]. 2016 [citado 9 de septiembre de 2022]. Disponible en: https://phecanada.ca/about.

59. Dudley D, Kriellaars D, Cairney J. Physical Literacy Assessment and Its Potential for Identification and Treatment of Children with Neuro-Developmental Behavioral Intellectual Disorders. Curr Dev Disord Reports. 2016;3(3):195-9.

60. Young L, O'Connor J, Alfrey L. Physical literacy: a concept analysis. Sport Educ Soc. 2020;25(8):946-59.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

61. Liu Y, Chen S. Physical literacy in children and adolescents: Definitions, assessments, and interventions. Eur Phys Educ Rev. 2021;27(1):96-112.

62. Mandigo J, Francis N, Lodewyk K, Lopez R. Physical Literacy for Educators. Phys Heal Educ J. 2009;75(3):27-30.

63. The Aspen Institute. Physical literacy: A global environmental scan [Internet]. Washington, DC; 2015. Disponible en: http://plreport.projectplay.us/.

64. Canada's Physical Literacy Consensus Statement [Internet]. 2015. p. 2. Disponible en: https://physicalliteracy.ca/physical-literacy/consensus-statement/.

65. Robinson DB, Randall L. Marking Physical Literacy or Missing the Mark on Physical Literacy? A Conceptual Critique of Canada's Physical Literacy Assessment Instruments. Meas Phys Educ Exerc Sci. 2017;21(1):40-55.

66. Young L, Alfrey L, O'Connor J. Moving from physical literacy to co-existing physical literacies: What is the problem? Eur Phys Educ Rev. 2022. In Press. Doi: 10.1177/1356336X221112867.

67. Keegan R, Barnett L, Dudley D. Physical Literacy: Informing a Definition and Standard for Australia. [Internet]. Australian Sports Comission; 2017. Disponible en: https://research-management.mq.edu.au/ws/portalfiles/portal/83466511/72163431.pdf.

68. Duncan MJ, Foweather L, Bardid F, Barnett AL, Rudd J, O'Brien W, et al. Motor Competence Among Children in the United Kingdom and Ireland: An Expert Statement on Behalf of the International Motor Development Research Consortium. J Mot Learn Dev. 2022;10(1):7-26.

69. Holfelder B, Schott N. Relationship of fundamental movement skills and physical activity in children and adolescents: A systematic review. Psychol Sport Exerc. 2014;15(4):382-91.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

70. Bolger LE, Bolger LA, O'Neill C, Coughlan E, O'Brien W, Lacey S, et al. Global levels of fundamental motor skills in children: A systematic review. J Sports Sci. 2021;39(7):717-53.

71. Goodway JD, Ozmun JC, Gallahue DL. Understanding Motor Development. Infants, Children, Adolescents, Adults. 8th Ed. Burlington, MA: Jones & Bartlett Learning; 2021.

72. Gallahue DL, Donnelly FC. Developmental physical education for all children. 4th Ed. Urban-Champaign, IL: Human Kinetics; 2007.

73. Corlett J, Mandigo J. A day in the life: Teaching physical literacy. Phys Heal Educ J. 2013;78(4):18-24.

74. Goodway JD, Robinson LE. SKIPing toward an Active Start Promoting Physical Activity in Preschoolers. Beyond J Young Child. 2006;61(3):1-6.

75. Barnett LM, Lai SK, Veldman SLC, Hardy LL, Cliff DP, Morgan PJ, et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. Sport Med. 2016;46(11):1663-88.

76. Barnett LM. The Routledge Handbook of Archaeothanatology. Brusseau TA, Fairclough SJ, Lubans DR, (Eds). The Routledge Handbook of Archaeothanatology. Routledge; 2020.

77. Butterfield SA, Angell RM, Mason CA. Age and sex differences in object control skills by children ages 5 to 14. Percept Mot Skills. 2012;114(1):261-74.

78. Young RW. The ontogeny of throwing and striking. Hum Ontog. 2009;3(1):19-31.

79. Gallahue DL, Ozmun JC, Goodway JD. Understanding motor development : Infants, children, adolescents, adults. 7th Ed. McGraw-Hill, New York: Humanities, Social Sciences & World Languages; 2012.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

80. Rudd JR, Barnett LM, Butson ML, Farrow D, Berry J, Polman RCJ. Fundamental movement skills are more than run, throw and catch: The role of stability skills. PLoS One. 2015;10(10):1-15.

81. Van Rossum T, Foweather L, Hayes S, Richardson D, Morley D. Expert recommendations for the design of a teacher-oriented movement assessment tool for children aged 4-7 years: a Delphi study. Meas Phys Educ Exerc Sci. 2021;25(4):283-93.

82. Logan SW, Barnett LM, Goodway JD, Stodden DF. Comparison of performance on process- and product-oriented assessments of fundamental motor skills across childhood. J Sports Sci. 2017;35(7):634-41.

83. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. J Sci Med Sport. 2019;22(3):311-8.

84. Hulteen RM, Barnett LM, True L, Lander NJ, del Pozo Cruz B, Lonsdale C. Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. J Sports Sci. 2020;1717-98.

85. Goodway JD, Robinson LE, Crowe H. Gender Differences in Fundamental Motor Skill Development in Disadvantaged Preschoolers From Two Geographical Regions. Res Q Exerc Sport. 2010;81(1):17-24.

86. Brian A, Pennell A, Taunton S, Starrett A, Howard-Shaughnessy C, Goodway JD, et al. Motor Competence Levels and Developmental Delay in Early Childhood: A Multicenter Cross-Sectional Study Conducted in the USA. Sports Med. 2019;49(10):1609-18.

87. Hardy LL, Barnett L, Espinel P, Okely AD. Thirteen-year trends in child and adolescent fundamental movement skills: 1997-2010. Med Sci Sports Exerc. 2013;45(10):1965-70.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

88. Barnett LM, Van Beurden E, Morgan PJ, Brooks LO, Beard JR. Does childhood motor skill proficiency predict adolescent fitness? Med Sci Sports Exerc. 2008;40(12):2137-44.

89. Lima RA, Pfeiffer K, Larsen LR, Bugge A, Moller NC, Anderson LB, et al. Physical activity and motor competence present a positive reciprocal longitudinal relationship across childhood and early adolescence. J Phys Act Heal. 2017;14(6):440-7.

90. Duncan MJ, Hall C, Eyre E, Barnett LM, James RS. Pre-schoolers fundamental movement skills predict BMI, physical activity, and sedentary behavior: A longitudinal study. Scand J Med Sci Sport. 2021;31(S1):8-14.

91. Bryant ES ara., James RS, Birch SL ouis., Duncan M. Prediction of habitual physical activity level and weight status from fundamental movement skill level. J Sports Sci. 2014;32(19):1775-82.

92. Jaakkola T, Yli-Piipari S, Huotari P, Watt A, Liukkonen J. Fundamental movement skills and physical fitness as predictors of physical activity: A 6-year follow-up study. Scand J Med Sci Sport. 2016;26(1):74-81.

93. Cattuzzo MT, dos Santos Henrique R, Ré AHN, de Oliveira IS, Melo BM, de Sousa Moura M, et al. Motor competence and health related physical fitness in youth: A systematic review. J Sci Med Sport. 2016;19(2):123-9.

94. Luz C, Rodrigues LP, De Meester A, Cordovil R. The relationship between motor competence and health-related fitness in children and adolescents. PLoS One. 2017;12(6):1-11.

95. Lopes VP, Stodden DF, Bianchi MM, Maia JAR, Rodrigues LP. Correlation between BMI and motor coordination in children. J Sci Med Sport. 2012;15(1):38-43.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

96. Babic MJ, Morgan PJ, Plotnikoff RC, Lonsdale C, White RL, Lubans DR. Physical Activity and Physical Self-Concept in Youth: Systematic Review and Meta-Analysis. Sport Med. 2014;44(11):1589-601.

97. Barnett LM, Morgan PJ, Van Beurden E, Ball K, Lubans DR. A reverse pathway? Actual and perceived skill proficiency and physical activity. Med Sci Sports Exerc. 2011;43(5):898-904.

98. Menescardi C, De Meester A, Morbée S, Haerens L, Estevan I. The role of motivation in the conceptual model of motor development in childhood. Psychol Sport Exerc. 2022;61: 102188.

99. Jaakkola T, Hillman C, Kalaja S, Liukkonen J. The associations among fundamental movement skills, self-reported physical activity and academic performance during junior high school in Finland. J Sports Sci. 2015;33(16):1719-29.

100. Haapala EA. Cardiorespiratory fitness and motor skills in relation to cognition and academic performance in children - A review. J Hum Kinet. 2013;36(1):55-68.

101. Leonard HC, Hill EL. Review: The impact of motor development on typical and atypical social cognition and language: A systematic review. Child Adolesc Ment Health. 2014;19(3):163-70.

102. O'Hagan AD, Behan S, Peers C, Belton S, O'Connor N, Issartel J. Do our movement skills impact our cognitive skills? Exploring the relationship between cognitive function and fundamental movement skills in primary school children. J Sci Med Sport. 2022. In Press. Doi: 10.1016/j.jsams.2022.08.001.

103. Dos Santos FG, Pacheco MM, Stodden D, Tani G, Maia JAR. Testing Seefeldt's Proficiency Barrier: A Longitudinal Study. Int J Environ Res Public Health. 2022;19(12):7184.

104. Cools W, Martelaer K de, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: A

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

review of seven movement skill assessment tools. J Sport Sci Med. 2009;8(2):154-68.

105. Eddy LH, Bingham DD, Crossley KL, Shahid NF, Ellingham-Khan M, Otteslev A, et al. The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. PLoS One. 2020;15(8):e0237919.

106. Klingberg B, Schranz N, Barnett LM, Booth V, Ferrar K. The feasibility of fundamental movement skill assessments for pre-school aged children. J Sports Sci. 2019;37(4):378-86.

107. Harris J, Cale L, Musson H. The predicament of primary physical education: A consequence of «insufficient» ITT and «ineffective» CPD? Phys Educ Sport Pedagog. 2014;17(4):367-81.

108. Van Rossum T, Foweather L, Richardson D, Hayes SJ, Morley D. Primary Teachers' Recommendations for the Development of a Teacher-Oriented Movement Assessment Tool for 4–7 Years Children. Meas Phys Educ Exerc Sci. 2019;23(2):124-34.

109. Eddy L, Hill LJB, Mon-Williams M, Preston N, Daly-Smith A, Medd G, et al. Fundamental Movement Skills and Their Assessment in Primary Schools from the Perspective of Teachers. Meas Phys Educ Exerc Sci. 2021;25(3):236-49.

110. Frankenburg W, Dodds J. The Denver developmental screening test. 1967;71(2):181-91.

111. Tabla de desarrollo Haizea– Llevant (0– 5 años). Gobierno de La Rioja. Consejería de Salud, Consumo y Bienestar Social.

112. Servicio Canario de Salud. Valoración del desarrollo psicomotor. (Programa de Salud Infantil) [Internet]. 2007 [citado 10 de septiembre de 2022]. Disponible en: https://www3.gobiernodecanarias.org/sanidad/scs/content/fb19

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5089-29f4-11e9-acfe-cba652c0f26c/ProgramaSaludInfantilCompleto.pdf.

113. Hoeboer J, De Vries S, Krijger-Hombergen M, Wormhoudt R, Drent A, Krabben K, et al. Validity of an Athletic Skills Track among 6- to 12-year-old children. J Sports Sci. 2016;34(21):2095-105.

114. Bruininks RH, Bruininks BD. Test of Motor Proficiency. 2nd edition. Manual. Circle Pines, Minnesota, USA: AGS Publishing; 2005.

115. Kiphard E, Schilling F. Körperkoordinationstest für Kinder KTK. Manual. 2. überarbeitete und ergänzte Auflage. 2nd Ed. Göttingen: Beltz Test; 2007.

116. Luz C, Rodrigues LP, Almeida G, Cordovil R. Development and validation of a model of motor competence in children and adolescents. J Sci Med Sport. 2016;19(7):568-72.

117. NSW Department of Education and Training. Get Skilled: Get Active. A K-6 resource to support the teaching of fundament skills. Sydney, New South Wales, Australia: NSW Department of Education and Training; 2016.

118. Ulrich DA. The standardization of a criterion-referenced test in fundamental motor and physical fitness skills. Diss Abstr Int. 1983;146A.

119. Ulrich DA. Test of Gross Motor Development - 2nd Ed. Austin, TX: Pro-Ed Publishers; 2000.

120. Department of Education Victoria. Fundamental Motor Skills. A Manual for Classroom Teachers [Internet]. Melbourne, Australia: Department of Education Melbourne; 1996. Disponible en: https://fusecontent.education.vic.gov.au/6f9265d9-afef-4a9a-9228-05a039225985/fmsteacher.pdf.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

121. Longmuir PE, Boyer C, Lloyd M, Borghese MM, Knight E, Saunders TJ, et al. Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age. J Sport Heal Sci. 2017;6(2):231-40.

122. López Pérez P, Cortés Psorio B. 17. Supervisión do desenvolvemento psicomotor. Guía de saúde infantil. [Internet]. Santiago de Compostela; 2020. Disponible en: https://www.sergas.es/Asistencia-sanitaria/Documents/1229/Guía saúde infantil_Cap. 17_Supervisión do desenvolvemento psicomotor.pdf.

123. Estevan I, Molina-García J, Bowe SJ, Álvarez O, Castillo I, Barnett LM. Who can best report on children's motor competence: Parents, teachers, or the children themselves? Psychol Sport Exerc. 2018;34:1-9.

124. Prieto JA, Cueto S, Carballo-Fazanes A, Abelairas-Gómez C. Psychomotor development disorders in apparently healthy children and considerations of family evaluation. J Hum Sport Exerc. 2022;17(3):518-27.

125. American Academy of Pediatrics. Early Childhood Visits. En: Hagan JF, Shaw JS, Duncan PM, (Eds.). Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescents [Internet]. 4th Ed. 2017. p. 501-647. Disponible en: https://brightfutures.aap.org/materials-and-tools/guidelines-and-pocket-guide/Pages/default.aspx.

126. Davies SC. Annual Report of the Chief Medical Officer 2012. Our Children Deserve Better: Prevention Pays [Internet]. Department of Health, editor. Our Children Deserve Better: Prevention Pays. 2013. 418 p. Disponible en: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/255237/2901304_CMO_complete_low_res_accessible.pdf.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

127. Ulrich DA. Test of Gross Motor Development. Austin, TX: Pro-Ed Publishers; 1985.

128. Ulrich DA. Test of Gross Motor Development- 3rd Ed. Austin, TX, USA: Pro-Ed Publishers; 2019.

129. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

130. Streiner DL, Norman GR, Cairney J. Health Measurement Scales a practical guide to their development and use. 5th Ed. Oxford, United Kingdom: OXFORD University Press; 2015.

131. DeVellis RF. Scale Development Theory and Applications. Fouth Edit. Paper Knowledge . Toward a Media History of Documents. Los Angeles: SAGE; 2017.

132. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sport Med. 1998;26(4):217-38.

133. Roberts P, Priest H. Reliability and validity in research. Nurs Stand. 2006;20(44):41-5.

134. Department of Educaction WA. Fundamental Movement Skills: Book 2. The Tools for Learning, Teaching and Assessment. The Government of Western Australia; 2013.

135. Robinson LE, Stodden DF, Barnett LM, Lopes VP, Logan SW, Rodrigues LP, et al. Motor Competence and its Effect on Positive Developmental Trajectories of Health. Sport Med. 2015;45(9):1273-84.

136. Valentini NC, Nobre GC, Zanella LW, Pereira KG, Albuquerque MR, Rudisill ME. Test of gross motor development-3 validity and reliability: A screening form. J Mot Learn Dev. 2021;9(3):438-55.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

137. Griffiths A, Toovey R, Morgan PE, Spittle AJ. Psychometric properties of gross motor assessment tools for children: A systematic review. BMJ Open. 2018;8(10):1-14.

138. Scheuer C, Herrmann C, Bund A. Motor tests for primary school aged children: A systematic review. J Sports Sci. 2019;37(10):1097-112.

139. Allen KA, Bredero B, Van Damme T, Ulrich DA, Simons J. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. J Autism Dev Disord. 2017;47(3):813-33.

140. Pan CY, Tsai CL, Chu CH. Fundamental Movement skills in children diagnosed with autism spectrum disorders and attention deficit hyperactivity disorder. J Autism Dev Disord. 2009;39(12):1694-705.

141. Houwen S, Hartman E, Jonker L, Visscher C. Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. Adapt Phys Act Q. 2010;27(2):143-59.

142. Magistro D, Piumatti G, Carlevaro F, Sherar L, Esliger D, Bardaglio G, et al. Measurement invariance of TGMD-3 in children with and without mental and behavioral disorders. Psychol Assess. 2018;30(11):1421-9.

143. Simons J, Daly D, Theodorou F, Caron C, Simons J, Andoniadou E. Validity and reliability of the TGMD-2 in 7-10-year-old flemish children with intellectual disability. Adapt Phys Act Q. 2008;25(1):71-82.

144. Rey E, Carballo-Fazanes A, Varela-Casal C, Abelairas-Gómez C. Reliability of the test of gross motor development: A systematic review. PLoS One. 2020;15:1-26.

145. Ayán C, Cancela JM, Sánchez-Lastra MA, Carballo-Roales AI, Domínguez-Meis F, Redondo-Gutiérrez L. Reliability and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

validity of the TGMD-2 battery in a Spanish population. Rev Iberoam Diagnostico y Eval Psicol. 2019;51(1):21-33.

146. Cano-Cappellacci M, Leyton FA, Carreño JD. Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica. 2015;49:97.

147. Farrokhi A, Zadeh ZM, Alvar KL, Kazemnejad A, Ilbeigi S. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City. J Phys Educ Sport Manag. 2014;5(2):18-28.

148. Kim S, Kim MJ, Valentini NC, Clark JE. Validity and reliability of the TGMD-2 for South Korean children. J Mot Behav. 2014;46(5):351-6.

149. Valentini NC, Barbosa MLL, Cini GV, Pick RK, Spessato BC, Balbinotti MAA. Teste de Desenvolvimento Motor Grosso: Validade e consistência interna para uma população Gaúcha. Rev Bras Cineantropometria e Desempenho Hum. 2008;10(4):399-404.

150. Valentini NC. Validity and Reliability of the TGMD-2 for Brazilian Children. J Mot Behav. 2012;44:275-80.

151. Lopes VP, Saraiva L, Rodrigues LP. Reliability and construct validity of the test of gross motor development-2 in Portuguese children. Int J Sport Exerc Psychol. 2016;16(3):250-60.

152. Mohammadi F, Bahram A, Khalaji H, Ulrich DA, Ghadiri F. Evaluation of the psychometric properties of the Persian version of the test of Gross Motor Development-3rd edition. J Mot Learn Dev. 2019;7(1):106-21.

153. Valentini NC, Zanell LW, Webster EK. Test of Gross Motor Development – Third Edition: Establishing Content and Construct Validity for Brazilian Children. J Mot Learn Dev. 2017;5(1):15-28.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

154. Webster EK, Ulrich DA. Evaluation of the Psychometric Properties of the Test of Gross Motor Development – 3rd Edition. J Mot Learn Dev. 2017;5(1):45-58.

155. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.

156. Lander N, Morgan PJ, Salmon J, Logan SW, Barnett LM. The reliability and validity of an authentic motor skill assessment tool for early adolescent girls in an Australian school setting. J Sci Med Sport. 2017;20(6):590-4.

157. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sport Med. 2008;38(4):297-316.

158. Aye T, Oo KS, Khin MT, Kuramoto-Ahuja T, Maruyama H. Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. J Phys Ther Sci. 2017;29(10):1726-31.

159. Capio CM, Eguia KF, Simons J. Test of gross motor development-2 for Filipino children with intellectual disability: Validity and reliability. J Sports Sci. 2016;34(1):10-7.

160. Estevan I, Molina-García J, Queralt A, álvarez O, Castillo I, Barnett L. Validity and reliability of the Spanish version of the Test of Gross Motor Development-3. J Mot Learn Dev. 2017;5(1):69-81.

161. Maeng H, Webster EK, Pitchford EA, Ulrich DA. Inter- and intrarater reliabilities of the test of gross motor development—third edition among experienced TGMD-2 raters. Adapt Phys Act Q. 2017;34(4):442-55.

162. Carballo-Fazanes A, Rey E, Valentini NC, Rodríguez-Fernández JE, Varela-casal C, Rico-Díaz J, et al. Intra-Rater (Live vs. Video Assessment) and Inter-Rater (Expert vs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Novice) Reliability of the Test of Gross Motor Development — Third Edition. Int J Environ Res Public Health. 2021;18:1652.

163. Rintala PO, Sääkslahti AK, Iivonen S. Reliability assessment of scores from video-recorded TGMD-3 performances. J Mot Learn Dev. 2017;5(1):59-68.

164. Valentini NC, Ramalho MH, Oliveira MA. Movement Assessment Battery for Children-2: Translation, reliability, and validity for Brazilian children. Res Dev Disabil. 2014;35(3):733-40.

165. Tyler R, Foweather L, Mackintosh KA, Stratton G. A dynamic assessment of children's physical competence: The dragon challenge. Med Sci Sports Exerc. 2018;50(12):2474-87.

166. Palmer KK, Brian A. Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. J Mot Learn Dev. 2016;4(2):142-51.

167. Cano-Cappellacci M, Leyton FA, Carreño JD. Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica. 2015;49:97.

168. Barnett LM, Minto C, Lander N, Hardy LL. Interrater reliability assessment using the Test of Gross Motor Development-2. J Sci Med Sport. 2014;17(6):667-70.

169. Kim Y, Park I, Kang M. Examining rater effects of the TGMD-2 on children with intellectual disability. Adapt Phys Act Q. 2012;29(4):346-65.

170. Kim C IL, Han DW, Park IH. Reliability and validity of the test of gross motor development-ii in Korean preschool children: Applying AHP. Res Dev Disabil. 2014;35(4):800-7.

171. Minick K, Kiesel K, Burton L, Taylor A, Plisky P, Butler R. Interrater Reliability of the Functional Movement Screen. Strength Cond. 2010;24(2):479-86.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

172. Gulgin H, Hoogenboom B. The functional movement screening (fms)^{TM}: an inter-rater reliability study between raters of varied experience. Int J Sports Phys Ther. 2014;9(1):14-20.

173. Kuo KT, Hunter BC, Obayashi M, Lider J, Teramoto M, Cortez M, et al. Novice vs expert inter-rater reliability of the balance error scoring system in children between the ages of 5 and 14. Gait Posture. 2021;86:13-6.

174. Onate J, Cortes N, Welch C, Van Lunen B. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. J Sport Rehabil. 2010;19(1):41-56.

175. Shultz R, Anderson SC, Matheson GO, Marcello B, Besier T. Test-retest and interrater reliability of the functional movement screen. J Athl Train. 2013;48(3):331-6.

176. Bandeira PFR, Duncan M, Pessoa ML, Soares Í, Da Silva L, Mota J, et al. Tgmd-2 short version: Evidence of validity and associations with sex, age, and bmi in preschool children. J Mot Learn Dev. 2020;8(3):528-43.

177. Valentini NC, Rudisill ME, Bandeira PFR, Hastie PA. The development of a short form of the Test of Gross Motor Development-2 in Brazilian children: Validity and reliability. Child Care Health Dev. 2018;44(5):759-65.

178. Duncan MJ, Martins C, Ribeiro Bandeira PF, Issartel J, Peers C, Belton S, et al. TGMD-3 short version: Evidence of validity and associations with sex in Irish children. J Sports Sci. 2022;40(2):138-45.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ANEXO 1. PUBLICACIONES

Publicación 1: *revisión sistemática* (apartado 6.1.1 de resultados)

Referencia: Rey E, **Carballo-Fazanes A**, Varela-Casal C, Abelairas-Gómez C; ALFA-MOV Project collaborators. Reliability of the test of gross motor development: A systematic review. PLoS ONE. 2020 16;15(7): e0236070.

Año de publicación: 2020

Doi: https://doi.org/10.1371/journal.pone.0236070

PMID: 32673358

Idioma: Inglés

Número de veces citado hasta octubre de 2022

- Web of Science: 14
- Scopus: 16
- Google Académico: 20

Datos extraídos del Journal Citation Reports (JCR) 2021	
Factor de Impacto	3.752
Factor de Impacto sin auto-citas	3.608
Factor de Impacto en 5 años:	4.069
Cuartil	2
Percentil	60.96
Categoría	Multidisciplinary sciences
Ranking	29/73

Autorización para el uso de la publicación: la política de Open Access de PloS ONE aplica la licencia Creative Commons Attribution

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(CC BY) a todos los trabajos publicados. Esto implica que los artículos estén disponibles para su reutilización sin permiso ni cuotas. Cualquier persona puede "*distribuir o reutilizar estos artículos, siempre que se cite adecuadamente al autor y a la fuente original*" (Información extraída de: *https://journals.plos.org/plosone/s/journal-information#loc-open-access*). Esta información aparece además en el propio artículo, bajo el título de "copyright".

Contribución de la doctoranda: La doctoranda participó en la conceptualización de la idea y diseño de la presente revisión sistemática. Por otra parte, participó en la fase de cribado y elegibilidad de los estudios incluidos en la revisión como tercera revisora cuando existió un desacuerdo entre los dos revisores principales. Además, fue una de las encargadas de la extracción de los datos de los artículos incluidos en la revisión sistemática y de la elaboración de las tablas del manuscrito que incluyen esta información resumida. Contribuyó en la redacción del borrador original así como en la revisión y edición de las siguientes versiones (conjuntamente con el resto de autores).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PLOS ONE

RESEARCH ARTICLE

Reliability of the test of gross motor development: A systematic review

Ezequiel Rey[1], Aida Carballo-Fazanes[2,3], Cristina Varela-Casal[1]*, Cristian Abelairas-Gómez[2,3,4]*, on behalf of ALFA-MOV Project collaborators[¶]

1 Faculty of Education and Sport Sciences, University of Vigo, Vigo, Spain, **2** Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain, **3** CLINURSID Research Group, Psychiatry, Radiology, Public Health, Nursing and Medicine Department, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, **4** Faculty of Education Sciences, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

¶ Membership of the ALFA-MOV Project is provided in the Acknowledgments.
* cristinavarelacasal@uvigo.es (CVC); cristianabelairasgomez@gmail.com (CAG)

Check for updates

OPEN ACCESS

Citation: Rey E, Carballo-Fazanes A, Varela-Casal C, Abelairas-Gómez C, on behalf of ALFA-MOV Project collaborators (2020) Reliability of the test of gross motor development: A systematic review. PLoS ONE 15(7): e0236070. https://doi.org/10.1371/journal.pone.0236070

Editor: Ali Montazeri, Iranian Institute for Health Sciences Research, ISLAMIC REPUBLIC OF IRAN

Received: April 10, 2020

Accepted: June 27, 2020

Published: July 16, 2020

Copyright: © 2020 Rey et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The ALFA-MOV Project is funded by FEDER/Spanish Ministry of Science, Innovation and Universities – National Agency of Investigation/ Project RTI2018-096106-A-I00.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Objective

To identify, synthesise and evaluate studies that investigated the reliability of the Test of Gross Motor Development (TGMD) variants.

Methods

A systematic search was employed to identify studies that have investigated internal consistency, inter-rater, intra-rater and test-retest reliability of the TGMD variants through Scopus, Pubmed/MEDLINE, PsycINFO, Sport Discus and Web of Science databases.

Results

Of the 265 studies identified, 23 were included. Internal consistency, evaluated in 14 studies, confirming good-to-excellent consistency for the overall score and general motor quotient (GMQ), and acceptable-to-excellent levels in both subscales (locomotor and ball skills). Inter-rater reliability, evaluated in 19 studies, showing good-to-excellent intra-class correlation coefficient (ICC) values in locomotor skills score, ball skills score, overall score, and GMQ. Intra-rater reliability, evaluated in 13 studies, displaying excellent ICC values in overall score and GMQ, and good-to-excellent ICC values in locomotor skills score and ball skills score. Test-retest reliability was evaluated in 15 studies with 100% of the statistics reported above the threshold of acceptable reliability when ICC was not used. Studies with ICC statistic showed good-to-excellent values in ball skills score, overall score, and GMQ; and moderate-to-excellent values in locomotor skills score.

Conclusions

Overall, the results of this systematic review indicate that, regardless of the variant of the test, the TMGD has moderate-to-excellent internal consistency, good-to-excellent inter-rater reliability, good-to-excellent intra-rater reliability, and moderate-to-excellent test-retest

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

reliability. Considering the few high-quality studies in terms of internal consistency, it would be recommend to carry out further studies in this field to improve their quality. Since there is no gold standard for assessing FMS, TGMD variants could be appropriate when opting for a psychometrical robust test. However, standardized training protocols for coding TGMD variants seem to be necessary both for researchers and practitioners in order to ensure acceptable reliability.

Introduction

Fundamental movement skills (FMS) are considered to be the "building blocks" for more developmentally advanced, complex movements essential for adequate participation in many organised and non-organised games, sports, or other specific physical activity [1–3]. FMS are typical classified into locomotor skills (e.g. running and hopping), manipulative or ball skills (e.g. catching and throwing), and stability skills (e.g. balancing and twisting) [1, 4]. Current evidence suggests that FMS competence is associated with better health outcomes in children and, in addition, this motor proficiency may have a potential role in promoting positive long-term health trajectories across the lifespan [5]. However, mastery of FMS does not emerge naturally [6], and learned exposure and environmental factors seems to play an important role in achieving a proficiency level in the period between early childhood (2–3 years) and later childhood (7–10 years) [7].

In light of previously reported health benefits, instruments used to assess and monitor motor proficiency have gained relevance in physical education over the last decades in order to identify students with motor deficiencies, to describe motor proficiency levels, and to support curricular decisions in schools [8]. FMS assessment tools can be broadly classified into two categories: quantity/product-oriented tests or quality/process-oriented tests [4, 9]. Product-oriented measures quantitatively assess the outcome of the movement (i.e. how far, how high) [10]. On the other hand, process-oriented assessment techniques evaluate the presence or absence of movement patterns demonstrated by a child providing qualitative information on children's motor competence that can be used for design and planning interventions [9, 11]. Among process-oriented assessment tools, the Test of Gross Motor Development (TGMD) and its variants (Test of Gross Motor Development–Second Edition [TGMD-2] and Test of Gross Motor Development–Third Edition [TGMD-3]) are, probably, the most frequently used technique for measuring FMS proficiency in educational, clinical, and research settings because of their low cost and feasibility [12–15]. The TGMD is a normative and criterion-based assessment designed to qualitatively evaluate the gross motor skill performance of children between the ages of 3 to 10 years and 11 months, with and without disabilities [13–15].

The TGMD is composed of two subscales, locomotor and object control/ball skills, which evaluate six to seven FMS with between three to five performance criteria, depending on skill [14, 15] (Table 1). Child performance is scored with 1 or 0 depending on the presence or absence of such criteria and the final raw scores can be converted into percentile ranks and standard scores. The test results can be used to identify children with gross motor developmental delay [16], to design, plan and evaluate the success of program interventions in FMS development, to assess individual progress, and to serve as an assessment tool in research [14].

Reliability can be considered a pre-requisite requirement for clinical, educational and research application of any given measure, even more for field-based measures, such as the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 1. TGMD-2 and TGMD-3 subscales, skills, performance criteria, and scores.

TGMD-2						TGMD-3					
Locomotor			Object Control			Locomotor			Ball Skills		
Skills	**Number of performance criteria**	**Max score**	**Skills**	**Number of performance criteria**	**Max score**	**Skills**	**Number of performance criteria**	**Max score**	**Skills**	**Number of performance criteria**	**Max score**
Run	4	8	Two hand strike: stationary	5	10	Run	4	8	Two hand strike: stationary	5	10
Gallop	4	8	Stationary dribble	4	8	Gallop	4	8	Forehand strike: self-bounced	4	8
Hop	5	10	Cath	3	6	Hop	4	8	Stationary dribble	3	6
Leap	3	6	Kick	4	8	Skip	3	6	Cath	3	6
Horizontal jump	4	8	Overhand throw	4	8	Horizontal jump	4	8	Kick	4	8
Slide	4	8	Underhand Roll	4	8	Slide	4	8	Overhand throw	4	8
									Underhand Roll	4	8

Max: Maximum.

https://doi.org/10.1371/journal.pone.0236070.t001

TGMD test. In this respect, in recent years, several studies have been published that examined the inter-rater, intra-rater, and test-retest reliability of the TGMD in different population groups, including children with autism spectrum disorder [17], children with attention deficit hyperactivity disorder [18], children with visual impairments [19], children with mental and behavioural disorders [20], and children with intellectual disabilities [21]. Given the increasing amount of scientific evidence on this topic and the extensive application of this assessment tool, a systematic review of the reliability of the TGMD appears to be warranted. Therefore, this study aimed to identify, synthesise and evaluate studies that investigated the reliability of the TGMD and critically appraise and summarise their results. The findings obtained may help clarify the true reliability of this test, and thus provide a valuable resource for practitioners and researchers interested in using the TGMD or interpreting its results.

Methods

Search strategy

This comprehensive systematic review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [22]. The searches were a combination of MeSH terms and free text words organised into three blocks: terms related to motor development, TGMD and reliability (S1 File). They were conducted through the following databases: Scopus, Pubmed/MEDLINE, PsycINFO, Sport Discus and Web of Science. Our PICO (Population, Intervention, Comparison, Outcomes) question [23] was as follows: Is the TGMD a reliable battery (O) in terms of internal consistency, inter-rater, intra-rater & test-retest reliability (C) to evaluate FMS (I) of pre- & schoolchildren (P)? The search was performed on 08 December, 2019.

Inclusion and exclusion criteria

The inclusion criteria were stablished in function of our PICO question:

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Participants. Studies with pre-school (≥3 & <6 years old) and/or schoolchildren (≥6 & <12 years old) participants were included. If data on age diverged between participants, only studies with ≤25% of the sample out of the range 3–11 years old were selected. If only statistics of age were reported, we selected those with mean age inside the range 3–11 and with Mean age + Standard Deviation ≤12. Studies omitting data on age were considered ineligible.

Intervention. We included articles in which FMS of pre- and/or schoolchildren were assessed with TGMD or any modified version (TGMD-2 / TGMD-3). Only research studies that embraced all the FMS of the TGMD or all the skills of one of both subscales (locomotor / ball skills) were included.

Comparison. We considered investigations that studied internal consistency, inter-rater, intra-rater and test-retest reliability of the TGMD.

Outcome. The main outcomes were a) internal consistency and b) inter-rater, intra-rater and test-retest reliability of locomotor, ball skills, overall and gross motor quotient (GMQ). Secondary outcomes were reliability assessment of each skill.

Type of study. We included original articles published in English, Spanish or Portuguese (from Portugal & Brazil). No minimum sample size was required.

Exclusion criteria. Studies whose principal aim was not to evaluate internal consistency, inter-rater, intra-rater and test-retest reliability of the TGMD were excluded. In this way, those manuscripts with secondary or additional results concerning reliability were not included. Commentary and opinion papers, abstracts, letters to editor, systematic reviews and meta-analysis were also excluded.

Study selection and data extraction

Both screening and eligibility were independently performed by two authors (E.R & C.A-G) to minimize potential bias. If there were disagreements, a third reviewer (A.C-F) was consulted to reach a decision. Data were independently extracted by two reviewers (A.C-F & C.A-G) based on minimum requirements recommended in the *Inclusion and exclusion criteria* [24], and were then cross-checked.

Methodological quality

Quality of the studies was evaluated using the COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) checklist following the COSMIN guideline for systematic reviews [25, 26], which includes 10 boxes with all standards needed to assess the quality of a study on different specific properties [25]. Boxes 4 and 6 were used in order to assess internal consistency and reliability, respectively. The COSMIN checklist evaluates *design requirements* (1 item for internal consistency & 3 items for reliability), *statistical methods* (1 item for both boxes) and the presence or not of *other important flaws in the design or statistical methods* (in both boxes). According to the COSMIN checklist, each item of each box is rated as *very good*, *adequate*, *doubtful* or *inadequate quality* [26]. The quality of each box corresponds with the lowest rating of any item of the box. The evaluation of risk of bias was appraised by two reviewers (A.C-F & C.A-G) using the tools available in COSMIN website (www.cosmin.nl). If there were disagreements and no consensus after discussion, a third reviewer (E.R) was consulted to reach a decision.

Manuscripts' statistics

Due to the large variety of statistical analyses observed in included studies, different reliability statistics classifications have been used. Internal consistency was assessed using Cronbach's alpha. According to the coefficient alpha size guidelines recommended by George and Malery

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[27], the following values were used to interpret Cronbach's alpha: $\alpha > 0.9$ –Excellent, $\alpha > 0.8$ –Good, $\alpha > 0.7$ –Acceptable, $\alpha > 0.6$ –Questionable, $\alpha > 0.5$ –Poor, and $\alpha < 0.5$ –Unacceptable. For inter-rater, intra-rater and test-retest reliability interpretation, a Pearson correlation > 0.80 [28] or ICC > 0.70 or Kappa > 0.70 [25, 28] was rated as "adequate reliability". Taking into account that ICC was the most used statistic in the included studies, to a more specific classification of reliability, the following ICC classification was used: ICCs less than 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and greater than 0.9 were classified as poor, moderate, good reliability, and excellent reliability, respectively [29]. Finally, for reliability analysis of each skill, 'adequate reliability' was operationally defined as ≥ 0.6 for ICCs, defined as the minimum useful level of agreement [30], sufficient for observing human movement for screening purposes [31].

Results

Summarize of studies

The initial search retrieved 238 abstracts and 27 additional studies were identified through other resources (i.e. by checking the list of references) (Fig 1). One-hundred and forty-two

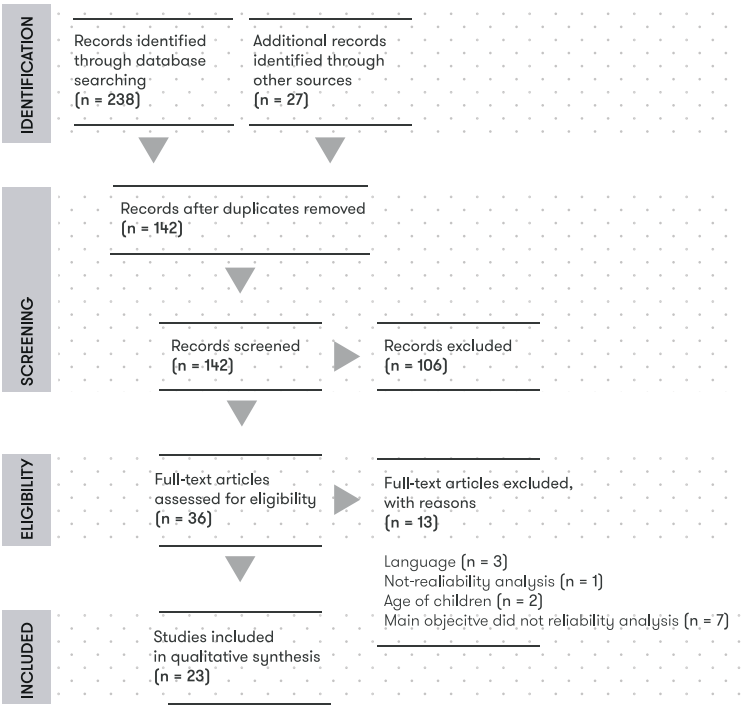


Fig 1. Flow diagram of the search and study selection process.

https://doi.org/10.1371/journal.pone.0236070.g001

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

abstracts were screened after removing duplicates and 23 studies were finally included. There was marked use of TGMD-2 [15 (65.2%)] *vs* TGMD-3 [8 (34.8%)]. Nineteen studies analysed locomotor and ball skills' score with overall score or GMQ, at least, in one of the reliability measurements (Table 2). Three studies only analysed locomotor and ball skills' score [32–34] and one only ball skills [35]. In most of cases, video recording was used for evaluating (n = 19). The studies were carried out in 14 different countries with participants aged between 4–9 years and around 40% were girls. Sample sizes ranged from 10 to 2674 participants. Table 3 shows the data extracted from the articles regarding internal consistency, inter-rater, intra-rater and test-retest reliability.

Internal consistency

Internal consistency was evaluated in 14 studies (8 TGMD-2 *vs* 6 TGMD-3) (Table 3). Alpha coefficients were calculated in different groups or participants for locomotor skills score (n = 18), ball skills score (n = 18), overall score (n = 12) and GMQ (n = 3). Ten studies reported alpha coefficients between 0.7 and 0.9 for locomotor skills score (4 over 0.9 [17, 36, 50, 51]), 11 for ball skills score (3 over 0.9 [36, 50, 51]), 4 for overall score (3 over 0.9 [17, 50, 51]) and 3 for GMQ (2 over 0.9 [21, 36]).

Inter-rater reliability

Inter-rater reliability was evaluated in 19 studies (12 TGMD-2 *vs* 7 TGMD-3). Three raters were used in 5 studies and 5 raters in 1 study. The rest of them (13) used 2 evaluators. Intra-class correlation (ICC) was calculated in most studies except in four in which Pearson correlation [21], Kappa statistic [33, 45] or content validity index (CVI) [38] were used. ≈70% of the inter-rater statistics reported regarding to locomotor and ball skills' score, overall score or GMQ were over 0.9. Only two studies shown inter-rater reliability values lower than 0.75 (ICC) [48] or 0.7 (kappa) [33], the last one comparing expert with novice coders. Scores of each individual skill were reported in 6 studies with more than ≈90% of the inter-rater reliability values over 0.6, ≈30% over 0.9 (ICC calculated in all comparisons) [35, 36, 42, 46, 48, 50] (Figs 2 and 3).

Intra-rater reliability

Intra-rater reliability was evaluated in 13 studies (6 TGMD-2 *vs* 7 TGMD-3). Two raters were used in 2 studies [39, 48] and 3 [49] and 5 [46] raters in 1 study. One study did not report the number of raters [40], and the remaining studies (8) used 2 evaluators. ICC and CVI were used in 12 and 1 study respectively. ≈85% of the intra-rater statistics shown regarding to locomotor and ball skills' score, overall score or GMQ were over 0.9, >95% over 0.75. Three studies reported specific data of each locomotor and ball skills, with ≈90% of intra-rater reliability values over 0.6, ≈35% over 0.9 (ICC calculated in all comparisons) [46, 48, 50] (Figs 4 and 5).

Test-retest reliability

Test-retest reliability was evaluated in 15 studies (10 TGMD-2 *vs* 5 TGMD-3). Test-retest reliability was evaluated using ICC (n = 6), Pearson correlation (n = 7), CVI (n = 1) and agreement ratio (n = 1). Reliability of TGMD measured over time showed values over 0.8 in 100% of the evaluations calculated with Pearson correlation, CVI and agreement ratio regarding to locomotor and ball skills' score, overall score or GMQ. In terms of ICC, more than 95% of values were over 0.75, 40% over 0.9. In three studies test-retest reliability was calculated for each

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

238

Table 2. Summary of the studies included in the review.

First author, year	Test	Country	Sample			Design	
			N & Profile	Sex	Age (in years)	Reliability assessment[a]	Viewing type
Allen, 2017 [17]	TGMD-3 traditional protocol	Australia	n = 14 children with ASD	Boys: 10 (71.4)	7.43 (2.03)	Internal consistency	Video evaluation
				Girls: 4 (28.6)		Inter-rater reliability	
	TGMD-3 visual support protocol		n = 21 typically developing children	Boys: 12 (57.1)	7.33 (1.75)	Intra-rater reliability	
				Girls: 9 (42.9)		Test-retest reliability	
Ayán, 2019 [36]	TGMD-2	Spain	n = 84 typically developing children	Boys: 46 (54.8)	8.35 (1.19)	Internal consistency	Video evaluation
				Girls: 38 (45.2)		Inter-rater reliability	
						Test-retest reliability	
Aye, 2017 [37]	TGMD-2	Myanmar	n = 50 typically developing children	Boys: 23 (46.0)	5.40 (0.30)	Inter-rater reliability	Video evaluation
				Girls: 27 (54.0)		Intra-rater reliability (n = 12)	
						Test-retest reliability (n = 25)	
Barnett, 2014 [35]	TGMD-2	Australia	n = 37 typically developing children	Boys: 13 (35.0)	6.20 (0.80)	Inter-rater reliability	Live evaluation
				Girls: 24 (65.0)			
Cano-Cappellacci, 2015 [38]	TGMD-2	Chile	n = 92 typically developing children	Boys: 56 (60.9)	7.50 (1.60)	Inter-rater reliability (n = 32)	Video evaluation
				Girls: 36 (39.1)		Intra-rater reliability (n = 38)	
						Test-retest reliability (n = 32)	
Capio, 2016 [39]	TGMD-2	Philippines	n = 81 children with intellectual disability	Boys: 65 (80.2)	9.29 (2.71)	Internal consistency	Video evaluation
				Girls: 16 (19.8)		Inter-rater reliability (n = 10)	
						Intra-rater reliability (n = 10)	
Estevan, 2017 [40]	TGMD-3	Spain	n = 178 typically developing children	Boys 93 (52.5)	6.94 (1.89)	Internal consistency	Video evaluation
				Girls 85 (47.5)		Inter-rater reliability (n = 4)	
						Intra-rater reliability (n = 4)	
Farrokhi, 2014 [41]	TGMD-2	Iran	n = 1438 typically developing children	Boys: 719 (50.0)	6.53 (2.25)	Internal consistency reliability	Video evaluation
				Girls: 719 (59.0)		Intra-rater reliability (n = 32)	
						Test-retest reliability (n = 63)	
Houwen, 2010 [19]	TGMD-2	Netherlands	n = 75 children with visual impairments	Boys: 46 (61.0)	8.50 (1.80)	Internal consistency	Video evaluation
				Girls: 29 (39.0)		Inter-rater reliability (n = 50)	
						Intra-rater reliability (n = 25)	
						Test-retest reliability (n = 23)	
Kim, 2012 [42]	TGMD-2	South Korea	n = 22 children with intellectual disability	Boys: 16 (72.7)	9.90 (1.30)	Inter-rater reliability	Video evaluation
				Girls: 6 (27.3)			
Kim S, 2014 [43]	TGMD-2	South Korea	n = 141 typically developing children	NR	6.80 (1.90)	Internal consistency reliability	Video evaluation
						Inter-rater reliability (n = 40)	
						Test-retest (n = 37)	

(*Continued*)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 2. (Continued)

First author, year	Test	Country	Sample			Design	
			N & Profile	Sex	Age (in years)	Reliability assessment[a]	Viewing type
Kim C-I, 2014 [44]	TGMD-2	South Korea	n = 121 typically developing children	Boys: 71 (58.7)	5.98 (0.32)	Internal consistency	NR
				Girls: 50 (41.3)			
Lopes, 2018 [45]	TMGD-2	Portugal	n = 330 typically developing children	Boys: 166 (50.3)	7.90 (1.30)	Internal consistency	Video evaluation
				Girls: 164 (49.7)		Inter-rater reliability	
						Test-retest reliability (n = 22)	
Maeng, 2017 [46]	TGMD-3	United States	n = 10 typically developing children	Boys: 6 (60.0)	6.57 (2.51)	Inter-rater reliability	Video evaluation
				Girls: 4 (40.0)		Intra-rater reliability	
Mohammadi, 2019 [47]	TGMD-3	Iran	n = 1600 typically developing children	Boys: 800 (50.0)	6.56 (2.29)	Internal consistency	Video evaluation
				Girls: 800 (50.0)		Inter-rater reliability (n = 160)	
						Intra-rater reliability (n = 160)	
						Test-retest reliability (n = 160)	
Palmer, 2016 [33]	TGMD-2	United States	n = 43 typically developing children	Boys: 25 (57.0)	4.88 (0.28)	Inter-rater reliability	Video evaluation
				Girls: 18 (43.0)			
Rintala, 2017 [48]	TGMD-3	Finland	n = 60 typically developing children	Boys: 28 (46.7)	3–7[b]	Inter-rater reliability (n = 20)	Video evaluation
				Girls: 32 (53.3)		Intra-rater reliability (n = 20 rater A and n = 20 rater B)	
Simons, 2008 [21]	TGMD-2	Belgium	n = 99 children with intellectual disability	Boys: 67 (67.7)	8.83 (1.75)	Internal consistency	Live evaluation
				Girls: 32 (32.3)		Inter-rater reliability	
						Test-retest reliability (n = 8)	
Valentini, 2008 [32]	TGMD-2	Brazil	n = 587 typically developing children	Boys: 300 (51.1)	7.52 (2.04)	Test-retest reliability	Video evaluation
				Girls: 287 (48.9)			
Valentini, 2012 [49]	TGMD-2	Brazil	n = 2674 typically developing children	Boys: 1352 (50.6)	7.56 (1.91)	Inter-rater reliability	Video evaluation
				Girls: 1322 (49.4)		Intra-rater reliability	
						Test-retest reliability (n = 648)	
Valentini, 2017 [50]	TGMD-3	Brazil	n = 597 typically developing children	Boys: 295 (49.4)	Boys: 6.76 (2.11)	Internal consistency reliability	Video evaluation
				Girls: 302 (50.6)	Girls: 6.58 (2.06)	Inter-rater reliability (n = 50)	
						Intra-rater reliability (n = 100)	
						Test-retest reliability (n = 128)	
Wagner, 2017 [34]	TGMD-3	Germany	n = 189 typically developing children	Boys: 99 (52.4)	7.15 (2.02)	Internal consistency	Video evaluation
				Girls: 90 (47.6)		Inter-rater reliability (n = 30)	
						Intra-rater reliability (n = 30)	
						Test-retest reliability (n = 104)	

(*Continued*)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 2. (Continued)

First author, year	Test	Country	Sample			Design	
			N & Profile	Sex	Age (in years)	Reliability assessment[a]	Viewing type
Webster, 2017 [51]	TGMD-3	United States	n = 807 typically developing children	Boys: 424 (52.5)	6.33 (2.09)	Internal consistency	Live and video evaluation
				Girls: 338 (47.5)		Test-retest reliability (n = 30)	

[a]: If sample used in each reliability assessment does not correspond with total sample size, it is specified in brackets.

[b]: Age shown as range. Mean and standard deviation of the whole sample size was not reportedsm Spectrum Disorder; *ICC*: Intra-class Correlation Coefficient; *NR*: not reported

Gender expressed as absolute frequencies (relative frequencies); *Age* expressed as mean (standard deviation)

https://doi.org/10.1371/journal.pone.0236070.t002

skill with reliability values over 0.8 in ≈50% (Pearson correlation calculated in all comparisons) [36, 49, 50] (Fig 6).

Children with disabilities

Five studies analysed some reliability measure in children with disabilities: autism syndrome disorder (ASD) (TGMD-3) [17], intellectual disability (TGMD-2) [21, 39, 42] and visual impairment (TGMD-2) [19]. In addition, in the case of ASD children, both protocols, traditional and with visual support were evaluated. Internal consistency was measured in four of the articles with values over 0.7 in all measurements regarding to locomotor and ball scores skills, overall score and GMQ. Inter-rater reliability was evaluated in the five manuscripts, while intra-rater and test-retest reliability were tested in three articles. High reliability was observed with scores over 0.9 in ≈90% of the cases in terms of inter- and intra-rater; ≈70% in test-retest.

Quality of studies

One study was classified as being of *very good* quality in terms of internal consistency and another one *insufficient*. The rest of the articles were classified as *doubtful*. The item of the COSMIN checklist with lower scores was the one referred to the calculation of statistics for each unidimensional scale or subscale separately. Inter-rater reliability was considered *very good* in 8 studies and *adequate* in other 8 (out of 19 studies). Similar results were found with regard to intra-rater reliability, with 5 studies with *very good* evaluation and 5 with *adequate* (out of 13 studies). Test-retest reliability was classified as *adequate* in 7 studies (out of 15 studies). More detailed results of the COSMIN quality evaluation is shown in Table 4.

Discussion

The TGMD is one of several process-oriented test batteries that purport to assess motor proficiency using visual observation in preschool and primary school-aged children [8, 52]. The purpose of this systematic review was to examine the literature related to the reliability of the TGMD and critically appraise and summarise their results. Generally, this review revealed strong psychometric properties for both TGMD-2 and TGMD-3, suggesting that TGMD variants could be a good choice when opting for a robust test in motor competence testing using product-oriented approaches.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 3. Main results.

First author, year	Test	Reliability results			
		Internal consistency	**Inter-rater reliability**	**Intra-rater reliability**	**Test-Retest**
Allen, 2017 [17]	TGMD-3	ASD Visual support protocol (n = 12):	ASD Visual support protocol (n = 12):	ASD Visual support protocol (n = 12):	ASD Visual support protocol (n = 8):
		LSS: α = 0.93; SEM = 3.28	LSS: ICC = 0.98, 95%CI (0.94–1.00); SEM = 1.75	LSS: ICC = 0.99, 95%CI (0.95–1.00); SEM = 1.24	LSS: ICC = 0.92, 95%CI (0.62–0.98); SEM = 2.99
		BSS: α = 0.81; SEM = 4.39	BSS: ICC = 0.96, 95%CI (0.86–0.99); SEM = 2.02	BSS: ICC = 1.00 95%CI (0.98–1.00); SEM = 0.71	BSS: ICC = 0.83, 95%CI (0.39–0.96); SEM = 2.83
		Overall: α = 0.93; SEM = 5.73	Overall: ICC = 0.99, 95%CI (0.95–1.00); SEM = 2.17	Overall: ICC = 0.99, 95%CI (0.98–1.00); SEM = 2.17	Overall: ICC = 0.92, 95%CI (0.66–0.98); SEM = 4.62
		ASD traditional protocol (n = 14):	ASD traditional protocol (n = 14):	ASD traditional protocol (n = 14):	ASD traditional protocol (n = 8):
		LSS: α = 0.82; SEM = 3.90	LSS: ICC = 0.98, 95%CI (0.92–0.99); SEM = 1.30	LSS: ICC = 0.97, 95%CI (0.88–0.99); SEM = 1.59	LSS: ICC = 0.92, 95%CI (0.65–0.98); SEM = 2.24
		BSS: α = 0.75; SEM = 4.63	BSS: ICC = 0.97, 95%CI (0.91–0.99); SEM = 1.60	BSS: ICC = 0.99, 95%CI (0.96–1.00); SEM = 0.93	BSS: ICC = 0.82, 95%CI (0.31–0.96); SEM = 2.93
		Overall: α = 0.88; SEM = 6.03	Overall: ICC = 0.98, 95%CI (0.94–1.00); SEM = 2.46	Overall: ICC = 0.99, 95%CI (0.92–1.00); SEM = 1.74	Overall: ICC = 0.91 95%CI (0.63–0.98); SEM = 4.06
		Typically developing (n = 21):	Typically developing (n = 21):	Typically developing (n = 21):	Typically developing (n = 17):
		LSS: α = 0.70; SEM = 3.20	LSS: ICC = 0.91, 95%CI (0.79–0.96); SEM = 1.76	LSS: ICC = 0.97, 95%CI (0.93–0.99); SEM = 1.01	LSS: ICC = 0.81, 95%CI (0.54–0.93); SEM = 2.22
		BSS: α = 0.60; SEM = 3.86	BSS: ICC = 0.92, 95%CI (0.81–0.97); SEM = 1.73	BSS: ICC = 0.91, 95%CI (0.68–0.97); SEM = 1.83	BSS: ICC = 0.84, 95%CI (0.62–0.94); SEM = 2.45
		Overall: α = 0.74; SEM = 5.18	Overall: ICC = 0.94, 95%CI (0.87–0.98); SEM = 2.49	Overall: ICC = 0.95, 95%CI (0.84–0.98); SEM = 2.27	Overall: ICC = 0.92, 95%CI (0.78–0.97); SEM = 2.72
Ayán, 2019 [36]	TGMD-2	LSS: α = 0.975	LSS: ICC = 0.976	NR	LSS: r = 0.952
		BSS: α = 0.963	BSS: ICC = 0.956		BSS: r = 0.929
		GMQ: α = 0.974	GMQ: ICC = 0.985		GMQ: r = 0.956
Aye, 2017 [37]	TGMD-2	NR	LSS: ICC = 0.95	LSS: ICC = 0.98	LSS: ICC = 0.82
			BSS: ICC = 0.88	BSS: ICC = 0.95	BSS: ICC = 0.79
			GMQ: ICC = 0.89	GMQ: ICC = 0.97	GMQ: ICC = 0.76
			Rater A x Rater B:		
			LSS: r = 0.97 (p<0.001)		
			BSS: r = 0.96 p<0.001)		
			GMQ: r = 0.97 (p<0.001)		
			Rater A x Rater C:		
			LSS: r = 0.94 (p<0.001)		
			BSS: r = 0.85 (p<0.001)		
			GMQ: r = 0.87 (p<0.001)		
			Rater B x Rater C:		
			LSS: r = 0.93 (p<0.001)		
			BSS: r = 0.96 p<0.001)		
Barnett, 2014 [35]	TGMD-2	NR	BSS: ICC = 0.93 95%CI (0.87–0.96)	NR	NR
Cano-Cappellacci, 2015 [38]	TGMD-2	NR	LSS: CVI = 0.87, 95%CI (0.73–0.93)	LSS: CVI = 0.92, 95%CI (0.83–0.95)	LSS: CVI = 0.86, 95%CI (0.71–0.93)
			BSS: CVI = 0.88 (0.77–0.94)	BSS: CVI = 0.86, 95%CI (0.76–0.93)	BSS: CVI = 0.80, 95%CI (0.59–0.90)
			Overall: CVI = 0.86, 95%CI (0.72–0.93)	Overall: CVI = 0.91, 95%CI (0.83–0.95)	Overall: CVI = 0.88, 95%CI (0.75–0.94)

(*Continued*)

USC 242 UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Anexo 1

Table 3. (Continued)

First author, year	Test	Reliability results			
		Internal consistency	**Inter-rater reliability**	**Intra-rater reliability**	**Test-Retest**
Capio, 2016 [39]	TGMD-2	LSS: α = 0.830	LSS: ICC = 0.996, 95%CI (0.984–0.999)	LSS: ICC = 0.995, 95%CI (0.978–0.999)	NR
		BSS: α = 0.792	BSS: ICC = 0.998, 95%CI (0.992–1.000)	BSS: ICC = 0.998, 95%CI (0.991–0.999)	
			Overall: ICC = 0.998, 95%CI (0.991–0.999)	Overall: ICC = 0.997, 95%CI (0.989–0.999)	
Estevan, 2017 [40]	TGMD-3	Locomotion: α = 0.80, 95%CI (0.75–0.84)	Overall: ICC = 0.90, 95%CI (0.66–0.98)	Overall: ICC = 0.98, 95%CI (0.85–1.00)	NR
		BSS: α = 0.85, 95%CI (0.81–0.88)			
		Overall: α = 0.89, 95%CI (0.87–0.92)			
Farrokhi, 2014 [41]	TGMD-2	LSS: α* = 0.78	NR	LSS: ICC = 0.95, 95%CI (0.91–0.97)	LSS: ICC = 0.65, 95%CI (0.50–0.79)
		BSS: α* = 0.74		BSS: ICC = 0.99 95%CI (0.97–0.99)	BSS: ICC = 0.85, 95%CI (0.75–0.91)
		GMQ: α* = 0.80		GMQ: ICC = 0.97, 95%CI (0.94–0.98)	GMQ: ICC = 0.81, 95%CI (0.70–0.89)
Houwen, 2010 [19]	TGMD-2	LSS: α = 0.71	LSS: ICC = 0.82, 95%CI (0.70–0.90)	LSS: ICC = 0.85, 95%CI (0.69–0.93)	LSS: ICC = 0.86, 95%CI (0.70–0.94)
		BSS: α = 0.72	BSS: ICC = 0.93, 95%CI (0.88–0.96)	BSS: ICC = 0.93, 95%CI (0.84–0.97)	BSS: ICC = 0.87, 95%CI (0.72–0.94)
			Overall: ICC = 0.89, 95%CI (0.81–0.93)	Overall: ICC = 0.95, 95%CI (0.88–0.98)	Overall: ICC = 0.92, 95%CI (0.82–0.91)
Kim, 2012 [42]	TGMD-2	NR	LSS: ICC = 0.80	NR	NR
			BSS: ICC = 0.75		
			Overall: ICC = 0.78		
Kim S, 2014 [43]	TGMD-2	LSS: α = 0.82	Rater A *vs* Rater B:	NR	LSS: r = 0.90
		BSS: α = 0.73	LSS: ICC = 0.94		BSS: r = 0.85
			BSS: ICC = 0.85		
		Overall: α = 0.87	Overall: ICC = 0.97		
			Rater A *vs* Rater C:		
			LSS: ICC = 0.90		
			BSS: ICC = 0.80		
			Overall: ICC = 0.93		
			Rater B *vs* Rater C:		
			LSS: ICC = 0.91		
			BSS: ICC = 0.77		
			Overall: ICC = 0.92		
			Rater A *vs* Rater B *vs* Rater C		
			LSS: ICC = 0.94		
			BSS: r = 0.92		
			Overall: ICC = 0.96		
Kim C-I, 2014 [44]	TGMD-2	Pre-weighting:	NR	NR	NR
		LSS: α = 0.53			
		BSS: α = 0.68			
		Overall: α = 0.72			
		Post-weighting			
		LSS: α = 0.66			
		BSS: α = 0.53			
		Overall: α = 0.70			

(*Continued*)

Table 3. (Continued)

First author, year	Test	Reliability results			
		Internal consistency	**Inter-rater reliability**	**Intra-rater reliability**	**Test-Retest**
Lopes, 2018 [45]	TMGD-2	LSS: α = 0.46	*Kappa statistic*	NR	*The Bland-Altman method*
		BSS: α = 0.64	The difference between test and retest measure varied between k = 0.7 (moderate consistency) & k = 1 (perfect consistency).		LSS:
		Overall: α = 0.69			95% of agreement ranged 0.85–1.17
					Agreement ratio: 1 (0.08)
					BSS:
					95% of agreement ranged 0.63–1.16
					Agreement ratio: 0.80 (0.13)
					Overall:
					95% of agreement ranged 0.80–1.13
					Agreement ratio: 0.96 (0.09)
Maeng, 2017 [46]	TGMD-3	NR	LSS: ICC = 0.92, 95%CI (0.82–0.98)	Rater A:	NR
			BSS: ICC = 0.93, 95%CI (0.84–0.98)	LSS: ICC = 0.99, 95%CI (0.98–0.99)	
			Overall: ICC = 0.96, 95%CI (0.91–0.99)	BSS: ICC = 0.97, 95%CI (0.90–0.99)	
				Overall: ICC = 0.99, 95%CI (0.95–0.99)	
				Rater B:	
				LSS: ICC = 0.97, 95%CI (0.90–0.99)	
				BSS: ICC = 0.98, 95%CI (0.91–0.99)	
				Overall: ICC = 0.98, 95%CI (0.91–0.99)	
				Rater C:	
				LSS: ICC = 0.99, 95%CI (0.95–0.99)	
				BSS: ICC = 0.97, 95%CI (0.87–0.99)	
				Overall: ICC = 0.98, 95%CI (0.93–0.99)	
				Rater D:	
				LSS: ICC = 0.94, 95%CI (0.76–0.98)	
				BSS: ICC = 0.93, 95%CI (0.73–0.98)	
				Overall: ICC = 0.95, 95%CI (0.81–0.99)	
				Rater E:	
				LSS: ICC = 0.99, 95%CI (0.98–0.99)	
				BSS: ICC = 0.99, 95%CI (0.98–0.99)	
				Overall: ICC = 0.99, 95%CI (0.98–0.99)	
				All raters:	
				LSS: ICC = 0.98, 95%CI (0.96–0.99)	
				BSS: ICC = 0.96, 95%CI (0.94–0.98)	
				Overall: ICC = 0.98, 95%CI (0.96–0.99)	

(*Continued*)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 3. (Continued)

First author, year	Test	Reliability results			
		Internal consistency	**Inter-rater reliability**	**Intra-rater reliability**	**Test-Retest**
Mohammadi, 2019 [47]	TGMD-3	LSS: α = 0.85	LSS: ICC = 0.97, 95%CI (0.96–0.98)	LSS: ICC = 0.98, 95%CI (0.98–0.99)	LSS: r = 0.92
		BSS: α = 0.85	BSS: ICC = 0.98, 95%CI (0.97–0.98)	BSS: ICC = 0.99, 95%CI (0.993–0.996)	BSS: r = 0.94
		Overall: α = 0.91	Overall: ICC = 0.98, 95%CI (0.97–0.98)	Overall: ICC = 0.99, 95%CI (0.994–0.997)	Overall: r = 0.95
Palmer, 2016 [33]	TGMD-2	NR	Novice vs expert coders	NR	NR
			LSS: k = -0.001		
			BSS: k = -0.004		
Rintala, 2017 [48]	TGMD-3	NR	LSS: ICC = 0.56	Rater A:	NR
			BSS: ICC = 0.64	LSS: ICC = 0.69	
			Overall: ICC = 0.62	BSS: ICC = 0.77	
				Overall: ICC = 0.75	
				Rater B:	
				LSS: ICC = 0.73	
				BSS: ICC = 0.73	
				Overall: ICC = 0.73	
Simons, 2008 [21]	TGMD-2	LSS: α = 0.82	LSS: r = 1.00	NR	LSS: r_s = 0.90
		BSS: α = 0.86	BSS: r = 1.00		BSS: r_s = 0.92
		GMQ: α = 0.90	GMQ: r = 1.00		GMQ: r_s = 0.98
Valentini, 2008 [32]	TGMD-2	NR	NR	NR	LSS: r = 0.82
					BSS: r = 0.88
Valentini, 2012 [49]	TGMD-2	NR	LSS: ICC = 0.88	ICC = 0.96	LSS: r = 0.83
			BSS: ICC = 0.89		BSS: r = 0.91
					Overall: r = 0.90
Valentini, 2017 [50]	TGMD-3	*Skills-to-test and subtests correlations*:	Rater 1 x Rater 2	LSS: ICC = 0.90	LSS: r = 0.93
		LSS: α = 0.63	LSS: ICC = 0.95	BSS: ICC = 0.85	BSS: r = 0.81
		BSS: α = 0.76	BSS: ICC = 0.97	Overall: ICC = 0.90	Overall: r = 0.90
		Overall: α = 0.74	Overall: ICC = 0.98		
		Performance-criteria-to-test and subtests correlations			
		LSS: α = 0.90			
		BSS: α = 0.88			
		Overall: α = 0.93			
Wagner, 2017 [34]	TGMD-3	LSS: α = 0.76	LSS: ICC = 0.88, 95%CI (0.76–0.95)	LSS: ICC = 0.97, 95%CI (0.94–0.99)	LSS: ICC = 0.94, 95%CI (0.91–0.96)
		BSS: α = 0.89	BSS: ICC = 0.97, 95%CI (0.94–0.99)	BSS: ICC = 0.99, 95%CI (0.98–1.00)	BSS: ICC = 0.98, 95%CI (0.97–0.99)
Webster, 2017 [51]	TGMD-3	LSS: α = 0.92	NR	NR	LSS: ICC = 0.97
		BSS: α = 0.95			BSS: ICC = 0.95
		Overall: α = 0.96			Overall: ICC = 0.97

TGMD: Test of gross motor development; *LSS*; Locomotor Skills Score; *BSS*; Ball Skills Score; *GMQ*: gross motor quotient; *ASD*: autism spectrum disorder; α: Cronbach's coefficient alpha; *SEM*: standard error of measurement; Overall: overall gross motor performance; *ICC*: Intra-class Correlation Coefficient; *CI*: Confidence interval; r: Pearson correlation coefficient; *NR*: Not reported; *CVI*: Content validity index; α*: Alpha no specified; k: kappa statistic; r_s: Spearman correlation.

https://doi.org/10.1371/journal.pone.0236070.t003

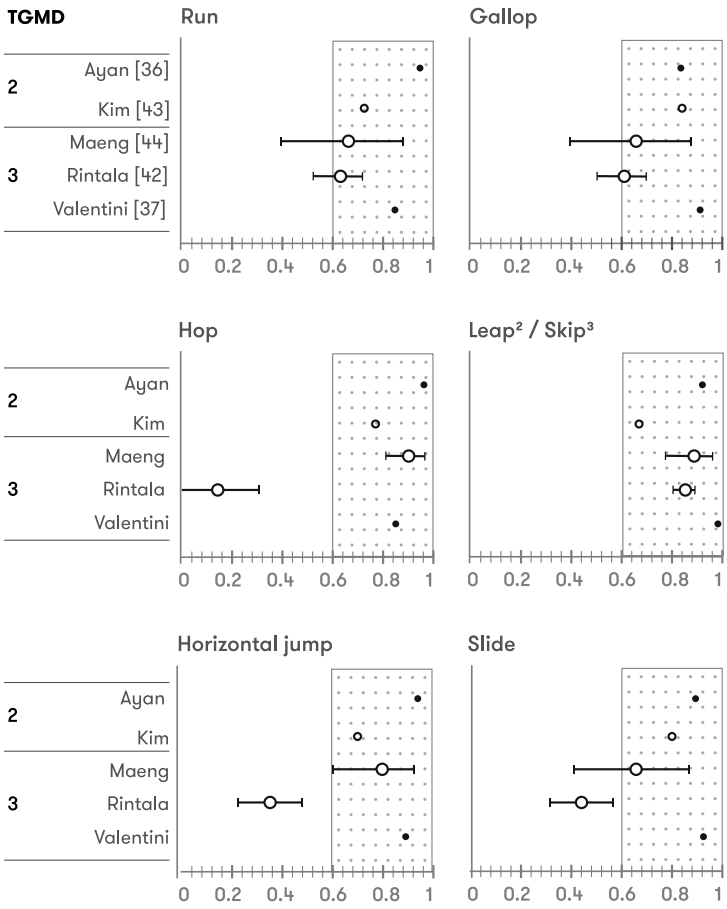


Fig 2. Inter-rater reliability of locomotor skills. *Open circles*: *very good* quality assessment according COSMIN checklist; *Closed circles*: *adequate* quality assessment according COSMIN checklist; *Larger circles*: ICC & 95%CI; *Smaller circles*: ICC.

https://doi.org/10.1371/journal.pone.0236070.g002

Internal consistency

Internal consistency refers to the degree to which test components (i.e. skills in TGMD variants) measure the same construct adequately (i.e. subscales and overall score in TGMD variants) [53]. The results from the 14 studies that evaluated internal consistency reliability confirmed, in most cases, good-to-excellent consistency for the TGMD-2 and TGMD-3 total score and GMQ, and acceptable-to-excellent levels of internal consistency in both subscales (locomotor and object

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

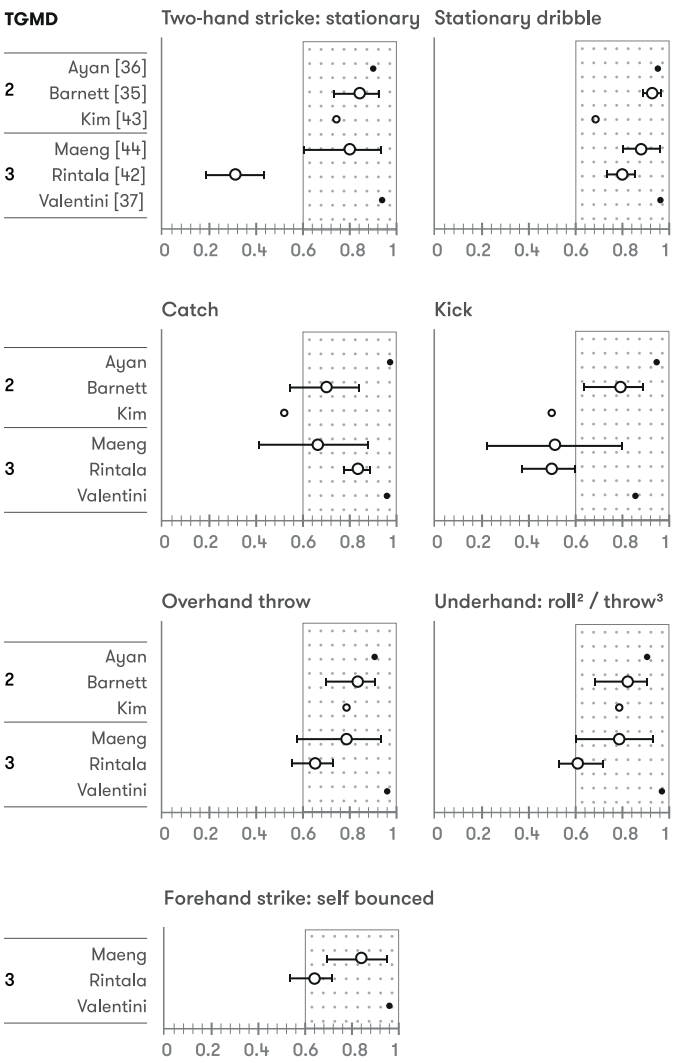


Fig 3. Inter-rater reliability of ball skills. *Open circles*: *very good* quality assessment according COSMIN checklist; *Closed circles*: *adequate* quality assessment according COSMIN checklist; *Larger circles*: ICC & 95%CI; *Smaller circles*: ICC.

https://doi.org/10.1371/journal.pone.0236070.g003

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

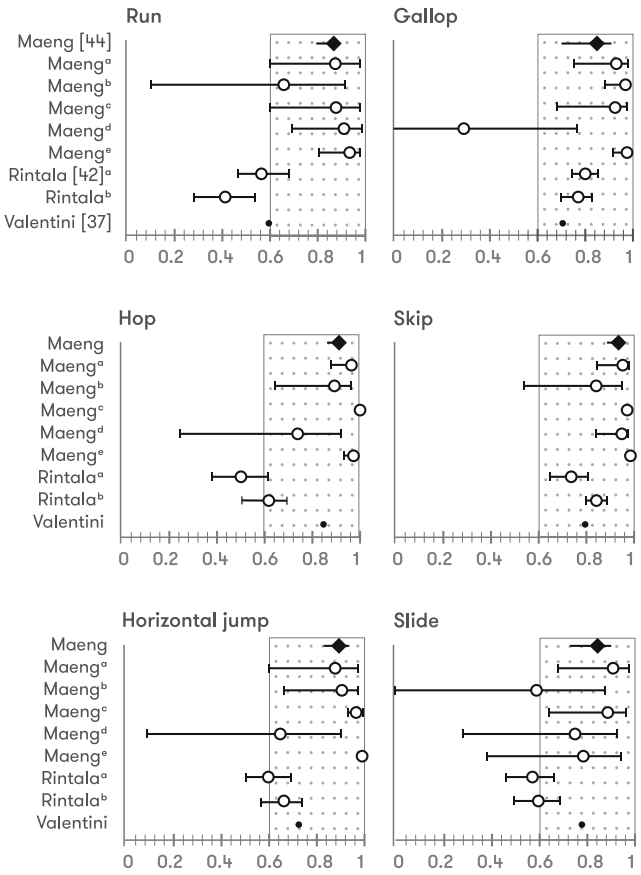


Fig 4. Intra-rater reliability of locomotor skills. [a-e]: Intra-rater reliability of each rater; *Diamonds*: Intra-rater reliability of all raters; *Open circles*: *very good* quality assessment according COSMIN checklist; *Closed circles*: *adequate* quality assessment according COSMIN checklist; *Larger circles*: ICC & 95%CI; *Smaller circles*: ICC.

https://doi.org/10.1371/journal.pone.0236070.g004

control/ball skills), indicating that the instrument seems to be consistent in evaluating the structures related to the subtests and total score in boys and girls [54]. In addition, skills and performance criteria seems to encompass a representation of the same construction [54].

Inter-rater reliability

Inter-rater reliability shows the agreement or consistency in scores from two or more raters, and is an essential psychometric property when assessing human movement skill proficiency

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

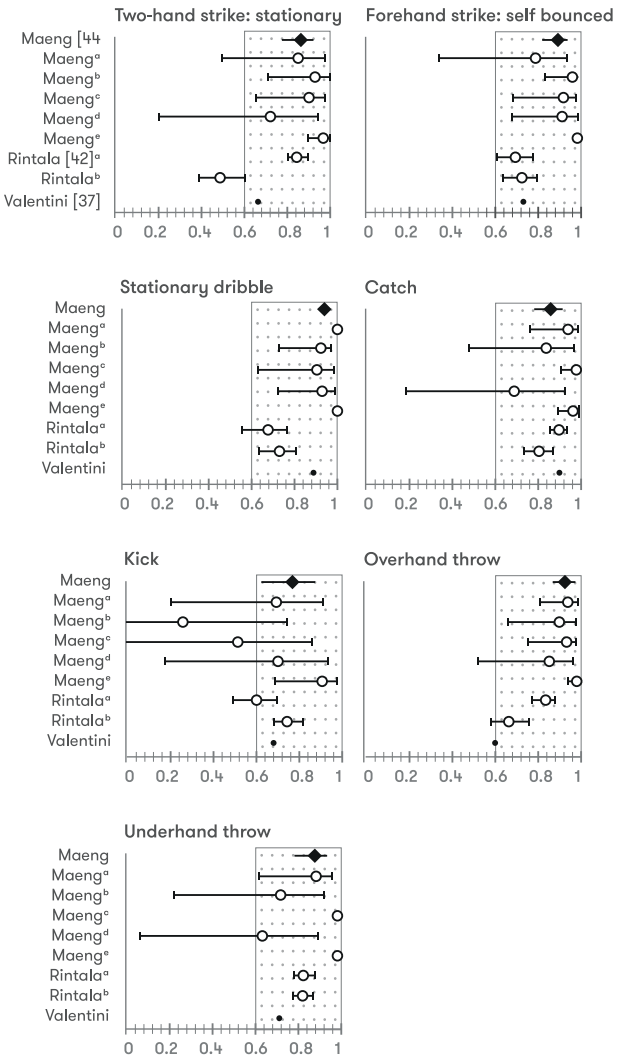


Fig 5. Intra-rater reliability of ball skills. [a-e]: Intra-rater reliability of each rater; *Diamonds*: Intra-rater reliability of all raters; *Open circles*: *very good* quality assessment according COSMIN checklist; *Closed circles*: *adequate* quality assessment according COSMIN checklist; *Larger circles*: ICC & 95%CI; *Smaller circles*: ICC.

https://doi.org/10.1371/journal.pone.0236070.g005

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

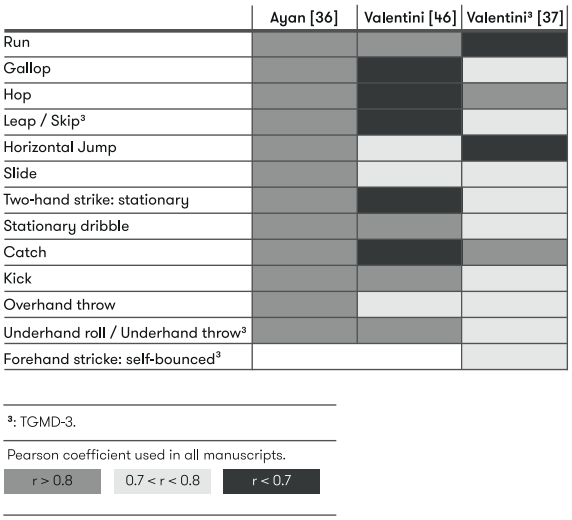


Fig 6. Test-retest reliability of locomotor and ball skills.

https://doi.org/10.1371/journal.pone.0236070.g006

[35]. The results from the 19 studies that evaluated inter-rated reliability confirmed, in most cases, adequate reliability levels and good-to-excellent ICC values for the TGMD-2 and TGMD-3 between raters in locomotor skills score, ball skills score, overall score, and GMQ, with ≈70% of the inter-rater statistics reported over 0.9 and 100% of coefficient values analysed above the defining thresholds of acceptable reliability for observing human movement screening. Only one study showed moderate levels of inter-rater reliability for locomotor and ball skills' score and overall score (in TGMD-3) [48]; primary due to the large variability observed among three individual skills (hop, horizontal jump, and two-hand strike). The inter-rater reliability values observed in this systematic review were similar to those reported in other product- and process-oriented instruments like Movement Assessment Battery for Children-2nd edition (MABC-2) [54], Bruininks-Oseretsky Test of Motor Proficiency–2nd Edition (BOT-2) [55], Basic Motor Competencies (MOBAK) [56], or Dragon Challenge [57].

Intra-rater reliability

Intra-rater reliability shows the degree of agreement among repeated evaluations of a test performed by the same rater. This review found excellent ICC values of intra-rater agreement for the TGMD variants in overall score and GMQ, and good-to-excellent in locomotor skills score and ball skills score. In addition, all but one [48] of the included studies reported adequate intra-rater reliability levels above the defining thresholds of acceptable reliability for this systematic review [25, 28]. Similar to inter-rater reliability analysis, only one study showed moderate levels of intra-rater reliability for locomotor skills score, ball skills score and overall score of TGMD-3, primarily due to the large variability observed among five individual skills (run, two-hand strike, slide, hop, and horizontal jump) [48]. Generally, the intra-rater reliability

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4. Quality assessment of the studies using the COSMIN checklist.

First author, year	Internal Consistency		Inter-rater	Intra-rater		Test-retest			
	Scales/ subscales	Statistics	Statistics	Time interval	Statistics	Patients stable	Time interval	Test conditions	Statistics
Allen, 2017 [17]	d	v	v	v	v	a	v	v	v
Ayán, 2019 [36]	v	v	a	--		a	v	v	d
Aye, 2017 [37]	--		v	v	v	a	v	a	v
Barnett, 2014 [35]	--		v	--		--			
Cano-Cappellacci, 2015 [38]	--		i	v	i	a	v	a	i
Capio, 2016 [39]	d	v	a	d	a	--			
Estevan, 2017 [40]	d	v	a	d	a	--			
Farrokhi, 2014 [41]	d	i	--	v	a	a	v	a	a
Houwen, 2010 [19]	d	v	v	v	v	a	v	v	v
Kim, 2012 [42]	--		v	--		--			
Kim S, 2014 [43]	d	v	a	--		a	v	a	d
Kim C-I, 2014 [44]	d	v	--	--		--			
Lopes, 2018 [45]	d	v	v	--		a	v	a	i
Maeng, 2017 [46]	--		v	v	v	--			
Mohammadi, 2019 [47]	d	v	a	v	a	a	v	a	d
Palmer, 2016 [33]	--		i	--		--			
Rintala, 2017 [48]	--		v	v	v	--			
Simons, 2008 [21]	d	v	d	--		a	v	a	d
Valentini, 2008 [32]	--		--	--		a	v	a	d
Valentini, 2012 [49]	--		a	d	a	v	v	a	a
Valentini, 2017 [50]	d	v	a	v	a	a	v	a	d
Wagner, 2016 [34]	d	v	a	v	a	a	v	a	a
Webster, 2017 [51]	d	v	--	--		a	v	a	a

v: Very good; *a*: Adequate; *d*: Doubtful; *i*: Insufficient.

https://doi.org/10.1371/journal.pone.0236070.t004

results of the studies included were somewhat higher than those observed for inter-rater reliability, supporting the evidence that is more likely that an evaluator will agree more consistently with him or herself than with other raters [19], which relates to the rater's subjectivity and discretion [38]. In order to minimise the probability that a rater would remember how he or she scored a specific child's performance from the previous scoring, the interval between evaluations is considered essential. Indeed, the time gap used in the studies included in this review to reduce memory-influenced bias varies from 12 days [41] to 3 months [48]. In addition, in three of these studies, the interval has not been specified [39, 40, 49]. Consequently, criteria relating to the time interval between tests used in the intra-rater reliability studies analysed seems to be due to an arbitrary chosen. Thus, further research should compare the intra-rater reliability of different TGMD variants using different time intervals to determine the optimal time gap to minimise memory bias.

Test-retest reliability

Test-retest reliability shows the temporal stability in scores measured by the same rater. Both TGMD variants revealed adequate levels of test-retests reliability, with 100% of the statistics reported above the defining thresholds of acceptable reliability for this systematic review [25, 28]. Specifically, TGMD-2 showed good-to-excellent ICC values of test-retest reliability (assessed in 10 studies) in overall score, GMQ, and ball skills score, and moderate-to-excellent

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ICC values of test-retest reliability in locomotor skills score. TGMD-3 (assessed in 5 studies) showed excellent ICC values of test-retest reliability in GMQ, and good-to-excellent ICC values in locomotor skills and ball skills. Test-retest reliability values observed in TGMD-2 and TGMD-3 were similar to those reported in other process-oriented instruments that assess individual skills in isolation, such as Victorian FMS Assessment [58].

Familiarisation of the evaluated participants with the testing procedures is an important factor that may influence reliability in a performance test [59]. In this regard, it is important to note that TGMD-2 and TGMD-3 examiner's manuals indicate that each participant should complete only one familiarisation trial for each skill after verbal description and demonstration of the evaluator [14, 15]. Thus, based on these results, test-retest reliability seems to be consistent regardless of the TGMD variant used and short familiarisation period.

Cultural and language adaptations

The different TGMD variants are widely used in several countries around the world. However, TGMD was developed for typically developing North American children. Due to the socio-cultural relevance of the subtests and the performance-criteria, several cross-cultural studies have investigated the psychometric properties of TGMD-2 and TGMD-3 in different languages, such as Spanish [36, 38, 40], Persian [47], German [34], or Portuguese [45, 49, 50] and/or cultures [43, 44]. Research conducted in this regard has described high and similar reliability characteristics to the original version, which evidences the clarity of TGMD instructions and the unambiguity of scoring [47].

Video-vs-live assessment

Although the TGMD examiner's manual does not assume videotaping assessment [48], most studies included in this review used video-recording evaluations (n = 19). TGMD videotaping evaluation seems to have several advantages as it allows more detailed scrutiny, assists observation of difficult performance criteria with slow-motion replay, and makes it possible to play each performance as many times as needed [48]. In addition, it is less time-consuming in educational settings as test scoring can be done outside classroom time. However, TGMD videotaping evaluation is not always possible due to different ethical considerations or the equipment required [35]. In this respect, it is important to note that the 3 studies that analysed TGMD reliability using live observation showed excellent values of inter-rater [21, 35], test-retest [21, 51], and good-to-excellent internal consistency [21, 51]. Intra-rater reliability was not assessed using live observation in any of the manuscripts included in this systematic review. Thus, TGMD variant reliability seems to be consistent regardless of the type of assessment. However, further research is needed to confirm these findings, comparing the reliability of different TGMD variants using live-versus-video assessment, and the association between rater training and the capacity to carry out live evaluation.

Rater training

According to the TGMD examiner's manual, supervised practice is recommended in administering and interpreting motor development tests, with at least three previous assessments before using TGMD in a real situation [14, 15]. However, rater familiarisation, training, and experience in TGMD administration were not systematically reported or described in the studies included in this systematic review. In addition, the academic background of the raters is heterogeneous, varying from graduate students (physical education [35, 36, 48] and sport sciences [36]), master's students [43, 44], doctoral students [33, 43, 50], physical therapists and physiatrists [37], or paediatric physiotherapists [39]. Previous evidence underscores the need

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

to provide standardised training protocols for coding using process-oriented approaches like TGMD-2 and TGMD-3 for valid and reliable results [46]. However, to the best of our knowledge, only one study analysed scoring differences using TGMD-2 between expert and novice coders [33]. The results showed that novice (undergraduate students in physical education with a two-hour training session on coding process) and expert (doctoral student in motor behaviour with more than 3 years of experience coding the TGMD-2) raters produced significantly different scores except for the kick and the gallop [33], suggesting a need for more extensive training until agreement is obtained. Thus, future research is necessary to explore the effects of providing standardised training protocols for coding TGMD-2 and TGMD-3 data and to determine the minimum training necessary to ensure acceptable reliability levels. In addition, future research should examine the subtest and the performance criteria in which the raters are mostly inconsistent, to paid special attention during familiarisation assessors.

Children with disabilities

While most of the studies included in this review have analysed reliability in typically developing children, five studies were conducted among children with intellectual disability [21, 39, 42], children with visual impairments [19], and children with ASD [17]. Generally, inter-rater (good-to-excellent), intra-rater (moderate-to-excellent) and test-retest (good-to-excellent), reliability values observed were similar to those reported in typically developing children. Based on these findings, TGMD variants could be considered an appropriate tool to examine FMS in these populations. However, the lower number of studies conducted in children with disabilities opens up an opportunity for future high-quality studies in these and other special populations.

Reliability of each skill

The reliability of each skill of TGMD-2 and TGMD-3 was evaluated in 7 studies [35, 36, 42, 50, 48, 46, 49]. In general, acceptable levels (ICC $\geq$ 0.6) of inter-rater reliability were observed for four locomotor skills (run, gallop, leap, and skip) and four ball skills (stationary dribble, overhand throw, underhand roll/throw, and forehand strike), showing moderate-to-excellent ICC values. However, the remaining three locomotor (hop, horizontal jump, and slide) and three ball skills (two-hand strike, catch, and kick) showed conflicting levels of inter-rater reliability. Differences in reliability between skills could be a reflection of the difficulty involved in assessing some skill components or performance criteria and the need to improve clarity in their scoring and interpretation.

Intra-rater reliability of individual skills were somewhat higher than those observed for inter-rater reliability, with seven skills (skip, horizontal jump, forehand strike, stationary dribble, catch, overhand throw, and underhand throw), showing moderate-to-excellent ICC values. The remaining six skills (run, gallop, hop, slide, two-hand strike, and kick) revealed conflicting levels of intra-rater reliability, which may reflect the need for more intensive training on the performance criteria evaluation for these specific skills [46]. It is important to note that the three studies that analysed intra-rater reliability of each skill used TGMD-3 version. Further research seems to be necessary to analyse this in TGMD-2, which is the most used variant of the test in scientific context.

Three studies evaluated test-retest reliability of each individual skill of TGMD-2 [36, 49] and TGMD-3) [50]. Several discrepancies were found in this regard in studies which used TGMD-2. Ayan et al [36] showed acceptable test-retest reliability levels (Pearson correlation $\geq$ 0.7) in all skills; however, Valentini [49] in seven skills (run, horizontal jump, slide, stationary dribble, kick, overhead throw and underhand roll). In the case of TGMD-3, only run and horizontal

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

jump were skills with low test-retest reliability values, which might reflect higher levels of temporal stability in TGMD-3 than TGMD-2 [50]. However, due to the low number of studies, to further explore this area, future research may be needed to confirm these findings.

Methodological quality

Fourteen studies evaluated internal consistency, and only one was classified as being of *very good* quality [36]. Any of the remaining studies did not calculated or expressed statistics for each unidimensional scales or subscales as it is highlighted [25]. That also involves calculation of Cronbach's alpha for each skill. In terms of inter-rater, intra-rater and test-retest reliability, the statistical methods item was the one which penalized the most. According with the COSMIN checklist, only using ICC (showing formula or model used) for continuous scores or kappa for dichotomous/nominal scores is possible to achieve a *very good* mark in this item [25]. However, previous studies suggested that coefficient of variation might be used in this regard with great applicability [60, 61]. Even so, most of manuscript in which inter- and intra-rater were evaluated were classified as being of *very good*/*adequate* quality.

Pearson correlation was used in the majority of the manuscripts in order to evaluate test-retest reliability. Nevertheless, this statistic is not considered the most suitable to assess reliability [25, 29, 62]. In addition, evidence that patients were stable between both evaluations is mandatory to be classified as *very good*. Since it might consider highly probable that children were stable during the evaluation, but no evidence was often provided, most manuscripts were classified in this item as *adequate*. Due to these rigorous and exigent methodological aspects in terms patients and statistics, no studies were classified as being of *very good* quality.

Limitations

A first limitation of this systematic review can be identified in the specific eligibility criteria that excluded the so-called grey literature. Thus, relevant publications could have been not included in this synthesis (i.e. monographs, conference abstracts, dissertations and theses). In addition, only publications in English, Spanish, or Portuguese that primary investigated reliability were selected. It can be assumed that significant articles could have been published in other languages. Another limitation was the absence of any form of meta-analysis in this systematic review due to the broad variety of statistical procedures employed to determine reliability and the heterogeneity of participants. Finally, as TGMD-3 variant is a relatively new test, the number of included studies that analysed psychometric properties of this version was significantly lower than TGMD-2.

Conclusions

A total of 23 studies were considered in this systematic review. Overall, the results of this systematic review indicate that, regardless of the variant of the test and the type of assessment (i.e. live vs. video), the TMGD has moderate-to-excellent internal consistency, good-to-excellent inter-rater reliability, good-to-excellent intra-rater reliability, and moderate-to-excellent test-retest reliability. Furthermore, reliability seems to be high both in typically developing children and children with disabilities; however, the lower number of studies in special populations reveals the need of further high-quality studies. Since there is no gold standard for assessing FMS, TGMD variants could be appropriate when opting for a psychometrical robust test. However, standardized training protocols for coding TGMD variants seem to be necessary both for researchers and practitioners in order to ensure acceptable reliability. Nevertheless, the optimal training protocol requires further study. Finally, due to the few high-quality

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

studies in terms of internal consistency, it would be recommend that further studies in this field refer to the COSMIN checklist to improve their quality.

Supporting information

S1 File. Research syntax.
(DOCX)

S1 Checklist.
(DOC)

Acknowledgments

We would like to thank the rest of the members of the ALFA-MOV Project; Barcala-Furelos R, Neiro-Rey C & Rodríguez-Fernández JE.

Author Contributions

Conceptualization: Ezequiel Rey, Aida Carballo-Fazanes, Cristian Abelairas-Gómez.

Data curation: Ezequiel Rey, Aida Carballo-Fazanes, Cristian Abelairas-Gómez.

Funding acquisition: Ezequiel Rey, Cristian Abelairas-Gómez.

Methodology: Ezequiel Rey, Aida Carballo-Fazanes, Cristian Abelairas-Gómez.

Project administration: Ezequiel Rey, Cristina Varela-Casal, Cristian Abelairas-Gómez.

Supervision: Cristian Abelairas-Gómez.

Visualization: Aida Carballo-Fazanes, Cristina Varela-Casal, Cristian Abelairas-Gómez.

Writing – original draft: Ezequiel Rey, Cristian Abelairas-Gómez.

Writing – review & editing: Ezequiel Rey, Aida Carballo-Fazanes, Cristina Varela-Casal, Cristian Abelairas-Gómez.

References

1. Gallahue D, Ozmun J, Goodway J. Understanding motor development: infants, children, adolescents, adults. 7th ed. New York, NY: McGraw-Hill; 2012.
2. Hulteen RM, Morgan PJ, Barnett LM, Stodden DF, Lubans DR. Development of Foundational Movement Skills: A Conceptual Model for Physical Activity Across the Lifespan. Sport Med. 2018; 48: 1533–40.
3. Logan SW, Ross SM, Chee K, Stodden DF, Robinson LE. Fundamental motor skills: A systematic review of terminology. J Sports Sci. 2018; 36: 781–96. https://doi.org/10.1080/02640414.2017.1340660 PMID: 28636423
4. Burton A, Miller D. Movement skill assessment. Champaing, IL: Human Kinetics; 1998.
5. Robinson LE, Stodden DF, Barnett LM, Lopes VP, Logan SW, Rodrigues LP, et al. Motor Competence and its Effect on Positive Developmental Trajectories of Health. Sport Med. 2015; 45: 1273–84.
6. Palmer KK, Chinn KM, Robinson LE. Using Achievement Goal Theory in Motor Skill Instruction: A Systematic Review. Sport Med. 2017; 47: 2569–83.
7. Tompsett C, Sanders R, Taylor C, Cobley S. Pedagogical Approaches to and Effects of Fundamental Movement Skill Interventions on Health Outcomes: A Systematic Review. Sport Med. 2017; 47: 1795–819.
8. Scheuer C, Herrmann C, Bund A. Motor tests for primary school aged children: A systematic review. J Sports Sci. 2019; 37: 1097–112. https://doi.org/10.1080/02640414.2018.1544535 PMID: 30604655
9. Yun J, Shapiro DR. A Quantitative Approach to Movement Skill Assessment for Children with Mental Retardation. Adapt Phys Act Q. 2004; 21: 269–80.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10. Barnett LM, Lai SK, Veldman SLC, Hardy LL, Cliff DP, Morgan PJ, et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. Sport Med. 2016; 46: 1663–88.
11. Bardid F, Vannozzi G, Logan SW, Hardy LL, Barnett LM. A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. J Sci Med Sport. 2019; 22: 311–8. https://doi.org/10.1016/j.jsams.2018.08.007 PMID: 30166086
12. Klingberg B, Schranz N, Barnett LM, Booth V, Ferrar K. The feasibility of fundamental movement skill assessments for pre-school aged children. J Sports Sci. 2019; 37: 378–86. https://doi.org/10.1080/02640414.2018.1504603 PMID: 30084306
13. Ulrich D. Test of Gross Motor Development. Austin, TX: Pro-Ed; 1985.
14. Ulrich DA. Test of Gross Motor Development (2nd ed.). Austin, TX: Pro-Ed; 2000.
15. Ulrich DA. Test of Gross Motor Development (3rd ed.). Austin, TX: Pro-Ed; 2016.
16. Brian A, Pennell A, Taunton S, Starrett A, Howard-Shaughnessy C, Goodway JD, et al. Motor Competence Levels and Developmental Delay in Early Childhood: A Multicenter Cross-Sectional Study Conducted in the USA. Sport Med. 2019; 49: 1609–18.
17. Allen KA, Bredero B, Van Damme T, Ulrich DA, Simons J. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. J Autism Dev Disord. 2017; 47: 813–33. https://doi.org/10.1007/s10803-016-3005-0 PMID: 28091840
18. Pan C-Y, Tsai C-L, Chu C-H. Fundamental Movement Skills in Children Diagnosed with Autism Spectrum Disorders and Attention Deficit Hyperactivity Disorder. J Autism Dev Disord. 2009; 39: 1694–705. https://doi.org/10.1007/s10803-009-0813-5 PMID: 19588236
19. Houwen S, Hartman E, Jonker L, Visscher C. Reliability and Validity of the TGMD-2 in Primary-School-Age Children With Visual Impairments. Adapt Phys Act Q. 2010; 27: 143–59.
20. Magistro D, Piumatti G, Carlevaro F, Sherar LB, Esliger DW, Bardaglio G, et al. Measurement invariance of TGMD-3 in children with and without mental and behavioral disorders. Psychol Assess. 2018; 30: 1421–9. https://doi.org/10.1037/pas0000587 PMID: 29792504
21. Simons J, Daly D, Theodorou F, Caron C, Simons J, Andoniadou E. Validity and Reliability of the TGMD-2 in 7–10-Year-Old Flemish Children with Intellectual Disability. Adapt Phys Act Q. 2008; 25: 71–82.
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339: b2535. https://doi.org/10.1136/bmj.b2535 PMID: 19622551
23. Richardson W, Wilson M, Nishikawa J, Hayward R. The well-built clinical question: a ~~key~~ to evidence-based decisions. ACP J Club. 1995; 123: A12–3.
24. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019) [Internet]. Cochrane. 2019.
25. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018; 27: 1171–9. https://doi.org/10.1007/s11136-017-1765-4 PMID: 29260445
26. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018; 27: 1147–57. https://doi.org/10.1007/s11136-018-1798-3 PMID: 29435801
27. George D, Mallery P. SPSS for Windows step by step: A Simple Guide and Reference, 11.0 update. 4th ed. Boston: Allyn & Bacon; 2003.
28. Chinapaw MJM, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB. Physical Activity Questionnaires for Youth: a systematic review of measuremen propierties. Sport Med. 2010; 40: 539–63.
29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016; 15: 155–63. https://doi.org/10.1016/j.jcm.2016.02.012 PMID: 27330520
30. Chinn S. Statistics in respiratory medicine. 2. Repeatability and method comparison. Thorax. 1991; 46: 454–6. https://doi.org/10.1136/thx.46.6.454 PMID: 1858087
31. Moran RW, Schneiders AG, Major KM, Sullivan SJ. How reliable are Functional Movement Screening scores? A systematic review of rater reliability. Br J Sports Med. 2016; 50: 527–36. https://doi.org/10.1136/bjsports-2015-094913 PMID: 26316583
32. Valentini N, Barbosa M, Cini G, Pick R, Spessate B, Balbinotti M. Teste de Desenvolvimento Motor Grosso: Validade e consistência interna para uma população Gaúcha. Rev Bras Cineantropometria Desempenho Hum. 2008; 10: 399–404.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

33. Palmer KK, Brian A. Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. J Mot Learn Dev. 2016; 4: 142–51.

34. Wagner MO, Webster EK, Ulrich DA. Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation): Results of a Pilot Study. J Mot Learn Dev. 2017; 5: 29–44.

35. Barnett LM, Minto C, Lander N, Hardy LL. Interrater reliability assessment using the Test of Gross Motor Development-2. J Sci Med Sport. 2014; 17: 667–70. https://doi.org/10.1016/j.jsams.2013.09.013 PMID: 24211133

36. Ayán C, Cancela J, Sánchez-Lastra M, Carballo-Roales A, Domínguez-Meis F, Redondo-Gutiérrez L. Fiabilidad y Validez de la Batería TGMD-2 en Población Española. Rev Iberoam Diagnóstico y Evaluación. 2019; 50: 21–34.

37. Aye T, Oo KS, Khin MT, Kuramoto-Ahuja T, Maruyama H. Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. J Phys Ther Sci. 2017; 29: 1726–31. https://doi.org/10.1589/jpts.29.1726 PMID: 29184278

38. Cano-Cappellacci M, Leyton FA, Carreño JD. Content validity and reliability of test of gross motor development in Chilean children. Rev Saude Publica. 2015; 49: 97.

39. Capio CM, Eguia KF, Simons J. Test of gross motor development-2 for Filipino children with intellectual disability: validity and reliability. J Sports Sci. 2016; 34: 10–7. https://doi.org/10.1080/02640414.2015.1033643 PMID: 25888083

40. Estevan I, Molina-García J, Queralt A, Álvarez O, Castillo I, Barnett L. Validity and Reliability of the Spanish Version of the Test of Gross Motor Development–3. J Mot Learn Dev. 2017; 5: 69–81.

41. Farrokhi A, Zareh Z, Karimi A, Kazemnejad A, Ilbeigi S. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3–10 aged children of Tehran City. J Phys Educ Sport Manag. 2014; 5: 18–28.

42. Kim Y, Park I, Kang M. Examining Rater Effects of the TGMD-2 on Children With Intellectual Disability. Adapt Phys Act Q. 2012; 29: 346–65.

43. Kim S, Kim MJ, Valentini NC, Clark JE. Validity and Reliability of the TGMD-2 for South Korean Children. J Mot Behav. 2014; 46: 351–6. https://doi.org/10.1080/00222895.2014.914886 PMID: 24915525

44. Kim C-I, Han D-W, Park I-H. Reliability and Validity of the Test of Gross Motor Development-II in Korean Preschool Children: Applying AHP. Res Dev Disabil. 2014; 35: 800–7. https://doi.org/10.1016/j.ridd.2014.01.019 PMID: 24529860

45. Lopes VP, Saraiva L, Rodrigues LP. Reliability and construct validity of the test of gross motor development-2 in Portuguese children. Int J Sport Exerc Psychol. 2018; 16: 250–60.

46. Maeng H, Webster EK, Pitchford EA, Ulrich DA. Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition Among Experienced TGMD-2 Raters. Adapt Phys Act Q. 2017; 34: 442–55.

47. Mohammadi F, Bahram A, Khalaji H, Ulrich DA, Ghadiri F. Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. J Mot Learn Dev. 2019; 7: 106–21.

48. Rintala PO, Sääkslahti AK, Iivonen S. Reliability Assessment of Scores From Video-Recorded TGMD-3 Performances. J Mot Learn Dev. 2017; 5: 59–68.

49. Valentini NC. Validity and Reliability of the TGMD-2 for Brazilian Children. J Mot Behav. 2012; 44: 275–80. https://doi.org/10.1080/00222895.2012.700967 PMID: 22857518

50. Valentini NC, Zanella LW, Webster EK. Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. J Mot Learn Dev. 2017; 5: 15–28.

51. Webster EK, Ulrich DA. Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition. J Mot Learn Dev. 2017; 5: 45–58.

52. Cools W, De Martelaer K, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med. 2009; 8: 154–68. PMID: 24149522

53. Thomas J, Nelson J. Research methods in physical activity. 2nd ed. Champaign, IL: Human Kinetics; 1990.

54. Valentini NC, Ramalho MH, Oliveira MA. Movement Assessment Battery for Children-2: Translation, reliability, and validity for Brazilian children. Res Dev Disabil. 2014; 35: 733–40. https://doi.org/10.1016/j.ridd.2013.10.028 PMID: 24290814

55. Bruininks R, Bruininks B. Bruininks-Oseretsky Test of Motor Proficiency–2nd Edition (BOT-2): Manual. Circle Pines, MN: AGS Publishing; 2005.

56. Scheuer C, Bund A, Becker W, Herrmann C. Development and validation of a survey instrument for detecting basic motor competencies in elementary school children. Cogent Educ. 2017; 4: 1337544.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

57. Tyler R, Foweather L, Mackintosh KA, Stratton G. A Dynamic Assessment of Children's Physical Competence. Med Sci Sport Exerc. 2018; 50: 2474–87.

58. Lander N, Morgan PJ, Salmon J, Logan SW, Barnett LM. The reliability and validity of an authentic motor skill assessment tool for early adolescent girls in an Australian school setting. J Sci Med Sport. 2017; 20: 590–4. https://doi.org/10.1016/j.jsams.2016.11.007 PMID: 28131506

59. Currell K, Jeukendrup AE. Validity, Reliability and Sensitivity of Measures of Sporting Performance. Sport Med. 2008; 38: 297–316.

60. Hopkins WG. Measures of Reliability in Sports Medicine and Science. Sport Med. 2000; 30: 1–15.

61. Grgic J, Oppici L, Mikulic P, Bangsbo J, Krustrup P, Pedisic Z. Test–Retest Reliability of the Yo-Yo Test: A Systematic Review. Sport Med. 2019; 49: 1547–57.

62. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60: 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012 PMID: 17161752

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Suplemento 1

Se empleó la siguiente estrategia de búsqueda:

Bloque 1: Motor development

"Fundamental motor skills" OR
"Fundamental motor skill proficiency" OR
"Fundamental movement skills" OR
"Fundamental movement skill proficiency" OR
"Motor skills" (MeSH)

AND

Bloque 2: TGMD

"Test of gross motor development" OR
tgmd

AND

Bloque 3: Tests & results
"Reliability and validity" (MeSH) OR
"Validity and reliability" (MeSH) OR
"Reproducibility of results" (MeSH) OR
"Reproducibility of findings" (MeSH) OR
Reliability OR
Validity OR
Reproducibility OR
Repeatability

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Publicación 2: *original* (apartado 6.1.2 de resultados)

Referencia: Carballo-Fazanes A, Rey E, Valentini NC, Rodríguez-Fernández JE, Varela-Casal C, Rico-Díaz J, Barcala-Furelos R, Abelairas-Gómez C. Intra-Rater (Live vs. Video Assessment) and Inter-Rater (Expert vs. Novice) Reliability of the Test of Gross Motor Development-Third Edition. Int J Environ Res Public Health. 2021; 18(4):1652.

Año de publicación: 2021

Doi: https://doi.org/10.3390/ijerph18041652

PMID: 33572298

Idioma: Inglés

Número de veces citado hasta octubre de 2022

- Web of Science: 2
- Scopus: 2
- Google Académico: 6

Datos extraídos del Journal Citation Reports (JCR) 2021	
Factor de Impacto	4.614
Factor de Impacto sin auto-citas	3.994
Factor de Impacto en 5 años:	4.708
Cuartil	1
Percentil	75.55
Categoría	Public, Environmental & occupational Health
Ranking	45/182

Autorización para el uso de la publicación: La política Open Access de la editorial MDPI siguiendo la licencia Creative Common CC BY de acceso abierto, permite la utilización libre de las publicaciones siempre y cuando se cite debidamente a la publicación original. Además, no se requieren permios especiales para reutilizar "*todo o*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

parte del artículo publicado en MDPI, incluidas las figuras y tablas" siempre que se cite claramente el artículo original (Información extraída de: https://www.mdpi.com/openaccess). Esta información aparece además en el propio artículo, bajo el título de "copyright".

Contribución de la doctoranda: La doctoranda contribuyó en el diseño del estudio y fue una de las encargadas de llevar a cabo las recogidas de datos asociadas, así como de la elaboración de las bases de datos necesarias para el análisis estadístico. Además, se encargó de la preparación y redacción del borrador original y también de las modificaciones posteriores con los cambios sugeridos por el resto de autores del artículo. Del mismo modo, al igual que el resto de autores, participó en la revisión del artículo siguiendo los comentarios y sugerencias de cambios de los revisores de la revista previos a la publicación.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

International Journal of
Environmental Research and Public Health

MDPI

Article

Intra-Rater (Live vs. Video Assessment) and Inter-Rater (Expert vs. Novice) Reliability of the Test of Gross Motor Development—Third Edition

Aida Carballo-Fazanes [1,2], Ezequiel Rey [3], Nadia C. Valentini [4], José E. Rodríguez-Fernández [5], Cristina Varela-Casal [3], Javier Rico-Díaz [5], Roberto Barcala-Furelos [1,3] and Cristian Abelairas-Gómez [1,2,5,*]

1 Health Research Institute of Santiago de Compostela (IDIS), University Hospital of Santiago de Compostela-CHUS, 15706 Santiago de Compostela, Spain; aidacarballofaz@gmail.com (A.C.-F.); roberto.barcala@uvigo.es (R.B.-F.)
2 CLINURSID Research Group, Psychiatry, Radiology, Public Health, Nursing and Medicine Department, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
3 REMOSS Research Group, Faculty of Education and Sport Sciences, Universidade de Vigo, 36005 Pontevedra, Spain; zequirey@uvigo.es (E.R.); cristinavarelacasal@uvigo.es (C.V.-C.)
4 Department of Physical Education, Physical Therapy and Dance, Universidade Federal do Rio Grande do Sul, Porto Alegre 90690-200, Brazil; nadiacv@esef.ufrgs.br
5 Faculty of Education Sciences, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain; geno.rodriguez@usc.es (J.E.R.-F.); javier.rico.diaz@usc.es (J.R.-D.)
* Correspondence: cristianabelairasgomez@gmail.com

check for updates

Citation: Carballo-Fazanes, A.; Rey, E.; Valentini, N.C.; Rodríguez-Fernández, J.E.; Varela-Casal, C.; Rico-Díaz, J.; Barcala-Furelos, R.; Abelairas-Gómez, C. Intra-Rater (Live vs. Video Assessment) and Inter-Rater (Expert vs. Novice) Reliability of the Test of Gross Motor Development—Third Edition. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1652. https://doi.org/10.3390/ijerph18041652

Academic Editor: Paul B. Tchounwou
Received: 28 December 2020
Accepted: 5 February 2021
Published: 9 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: The Test of Gross Motor Development (TGMD) is one of the most common tools for assessing the fundamental movement skills (FMS) in children between 3 and 10 years. This study aimed to examine the intra-rater and inter-rater reliability of the TGMD—3rd Edition (TGMD-3) between expert and novice raters using live and video assessment. Five raters [2 experts and 3 novices (one of them BSc in Physical Education and Sport Science)] assessed and scored the performance of the TGMD-3 of 25 healthy children [Female: 60%; mean (standard deviation) age 9.16 (1.31)]. Schoolchildren were attending at one public elementary school during the academic year 2019–2020 from Santiago de Compostela (Spain). Raters scored each children performance through two viewing moods (live and slow-motion). The ICC (Intraclass Correlation Coefficient) was used to determine the agreement between raters. Our results showed moderate-to-excellent intra-rater reliability for overall score and locomotor and ball skills subscales; moderate-to-good inter-rater reliability for overall and ball skills; and poor-to-good for locomotor subscale. Higher intra-rater reliability was achieved by the expert raters and novice rater with physical education background compared to novice raters. However, the inter-rater reliability was more variable in all the raters regardless of their experience or background. No significant differences in reliability were found when comparing live and video assessments. For clinical practice, it would be recommended that raters reach an agreement before the assessment to avoid subjective interpretations that might distort the results.

Keywords: TGMD-3; agreement; assessment; fundamental motor skills; raters; live; slow-motion

1. Introduction

Fundamental movements skills (FMS) consist of a basic organized movement involving the combination of movement patterns of two or more parts of the body [1]. FMS are considered "building blocks" for more advanced and complex movements necessary to participate in different sports, games and other physical activities. Commonly, FMS are classified into locomotor skills (e.g., run, jump, hop, slide), object control/ball skills (e.g., catch, kick, strike, throw) and balance/stability skills (e.g., static balance, dynamic balance) [1–3].

Strong existing evidence suggests positive associations between FMS competency and physical activity [4], physical fitness [4,5] and health-related benefits [6,7] such as

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

healthy weight status [4,5], cardiorespiratory fitness [4,5] or positive cognitive and academic outcomes [8] among others. However, the acquisition of FMS proficiency levels does not occur naturally and the implementation of structured physical education programs in pre-school and school-aged children is required [9].

In light of the aforementioned reported benefits, existing literature supports the importance of assessing FMS using valid and reliable tools [10], which is also necessary to provide valuable information about children's motor performance and progress [11]. It is essential to early identification of possible delays or disorders that could affect motor competence and cognitive [12] and affective [13] development.

There are several assessment tools to evaluate FMS, which can be broadly classified into two categories: quantity/product-oriented assessment, which evaluates the outcome of the movement (e.g., velocity, trajectory) and quality/process-oriented assessment, which evaluates the pattern of the movement (e.g., how a person throws) [14,15]. Moreover, assessments may examine gross motor skills, movements that require the use of large motor groups (e.g., running, jumping) [16] or fine motor skills, movements that involve small motor groups (writing, eating) [17].

One of the most common assessments used to examine FMS in children is the Test of Gross Motor Development (TGMD) [3] and its variants TGMD—Second edition (TGMD-2) [18] and TGMD—Third Edition (TGMD-3) [19]. The TGMD-3 is a validated and reliable [20] process-oriented assessment applied to evaluate gross motor competence in children between 3 years-0 months and 10 years-11 months. The TGMD-3 assess thirteen FMS divide into two subscales, locomotor and ball skills.

Reliability is one of the most essential and fundamental features in assessing performance in research. For one hand, reliability is an essential issue in research, since it lets researchers replicate studies. On the other hand, from a practical perspective, it would be important to assess accurately those variables that wanted to be studied. For example, regarding the scope of the present study, the evaluation of FMS by schoolteachers or clinical staff. In this sense, a current systematic review [20] of 23 studies assessing the reliability of the TGMD showed good-to-excellent inter-rater reliability, good-to-excellent intra-rater reliability and moderate-to-excellent test-retest reliability. Most studies assessed the reliability among experienced raters using video evaluation. However, the TGMD is sometimes used by researchers or practitioners (such as physical education teachers) with little or no training [21], since the examiner's manual only recommend to train with at least three children before a diagnostic evaluation [18,19]. Thus, it is necessary to know reliability among raters with different backgrounds and experiences. Only one previous study has compared differences in scores between expert and novice coders [22]. The results indicated that novice raters could not score the TGMD-2 in a significantly similar manner to the experts [22]. However, to this day, no study examined this issue in TGMD-3.

Although the TGMD examiner's manual does not refer to the children's assessment through video recordings, many studies used videotaped assessment because of its advantages over live evaluation. It allows the evaluation of each skill's criteria in more detail and repeatedly, even in slow-motion if necessary [22]. Intra and inter-rater reliability between expert and novice raters considering a different type of video viewing, live and slow-motion, has not yet been investigated. Thus, the purpose of this study was to analyze intra- and inter-rater reliability of TGMD-3 using live and video assessments between experts and novice raters.

2. Materials and Methods

2.1. Participants

This study was conducted with 25 healthy children (60% females, $n = 15$; 100% Hispanic, $n = 25$) between 8 and 10 years old (Mean $\pm$ SD: 9.16 $\pm$ SD 1.31 years), attending a public elementary school during the academic year 2019-2020 in Santiago de Compostela, Galicia, Spain. Participants' anthropometric were obtained for each child (height: 1.37 $\pm$ 0.11 m; weight: 34.57 $\pm$ 8.49 kg; and body mass index: 18.26 $\pm$ 2.95 kg·m^{-2}).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Participation in the study was voluntary and previously, parents or guardians signed the informed consent. All children were provided with verbal information and gave verbal assent before the test. This study respected the Helsinki Convention's ethical principles and was approved by the Ethical Committee of the University of Vigo, Spain.

2.2. Raters

Five raters (convenience sample), two experts (ER) and three novices (NR) were responsible for assessing the participants' video recorders. The three NR had no previous experience rating TGMD-3. One of the novices had a BSc in Physical Education and Sports Science (NR-PE); the other two were a primary schoolteacher and a nurse, both with no physical education or sports science background and experience in working with children either. However, all of them might have to assess FMS of children in the future due to their academic/professional background. Before coding children's performance, NR studied the content of the TGMD manual [19], enquired the expert coders about their doubts and practiced administering the TGMD-3 to three children. The two ER were BSc in Physical Education and Sports Science. Both with more than 5 years of experience testing the TGMD.

2.3. Test of Gross Motor Development—Third Edition [TGMD-3]

The TGMD-3 assesses gross motor skill performance in children between 3 years and 0 months and 11 years and 11 months. It consists of two subscales: locomotor and ball skills. The locomotor subscale measures the gross motor skills that involve fluid coordination of the body while the child moves in one direction or another and includes: run, gallop, hop, skip, horizontal jump and slide. The ball skills subscale assesses the gross motor skills that require effectiveness in intercepting and propelling objects and involves: striking a stationary ball, forehand strike, stationary dribble, catch, kick, overhand throw and underhand throw. Each skill, which includes several behavioral components, was evaluated through three to five criteria, scored 1 or 0 depending on the presence or absence [19]. We obtain independent scores for each skill, subscale and total test from the criteria scores.

After a verbal description and a practical demonstration, the children perform three trials of each skill. The first trial is a practice and it was not scored. The examiner scores the remaining two trials as follows: if the child performs a behavioral component or criterion correctly, the examiner scores a 1 and if the child does not perform it correctly, a 0. According to this, the maximum score that each child can achieve is 100 points (46 for the locomotor subscale and 54 for the ball skills subscale).

2.4. Procedures

A physical education teacher and a nurse (PhD student) conducted the assessments. All data were collected in March 2020. The tests were carried out in a sports hall, during the school's regular schedule and always in the presence of at least one teacher at the school. The testers provided the children with a verbal description of each skill, followed by a video showing the skill's correct performance (showed twice: normal speed and slow-motion) following the TGMD-3 guidelines [19]. Each child performed a practice trial followed by two consecutive trials, which were video-recorded (camera Nikon D5300) for later reliability analysis. The administration of the test to each child took approximately 20–30 min.

The five raters independently assessed the recorded videos to analyze the inter- and intra-rater reliability. Raters assessed the video in two viewing modes: live (once, at normal speed and without pause) and slow-motion (watching the video as many times as they needed, in slow-motion and with pauses). The children, the order of skills and the type of visualization (live/slow-motion) for each video was randomized for each rater. For the study of intra-rater reliability, each rater re-assessed the videos following the corresponding viewing mode according to the randomization after 2 weeks. In this sense, intra-rater reliability was analyzed according to the viewing mode (live vs. slow-motion)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and the inter-rater reliability was studied between NR and ER in both conditions, live and slow-motion.

2.5. Statistical Analysis

Reliability was assessed by intraclass correlation coefficients (ICC) for both intra- and inter-rater. The selection process of ICC was based on Koo et al. [23] flow-chart. ICC values and their 95% confident intervals for intra-rater reliability were based on single-measurement, absolute-agreement, 2-way mixed-effects model. In the case of inter-rater reliability, single-measurement, consistency, 2-way random-effects model. Values less than 0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability and values greater than 0.90 indicate excellent reliability.

Additionally, repeated-measures ANOVA was performed to analyze the overall scores and the locomotor and ball skills subscales' scores; means and standard deviations were provided. Two factors were integrated into the model: one intra-group (live vs. slow-motion) and one inter-group (raters). Raters were categorized as follows: NR (Rater A and Rater B), NR-PE (Rater C) and ER (Rater D and Rater E).

All analyses were performed using the SPSS statistical package version 23 (SPSS Inc, Chicago, IL). A significance level of $p < 0.05$ was considered.

3. Results

3.1. Intra-Rater Reliability: Live vs. Slow-Motion

Results of intra-rater reliability are shown in Table 1. Good-to-excellent or excellent reliability were found for most of the skills, mainly in ER and NR-PE evaluations. ICC values from subscales' and overall scores also showed good-to-excellent or excellent reliability in the case of ER and NR-PE. NR showed moderate-to-excellent, good-to-excellent or excellent intra-rater reliability in both subscales' and overall scores.

ER and NR-PE achieved intra-rater reliability at least moderate in all skills (except ER (Rater E): slide). In the case of NR, run (Rater A and B), hop (Rater B), horizontal jump (Rater A and B), kick (Rater A) and underhand throw (Rater B) were the skills in which reliability was poor-to-moderate/good.

Descriptive statistics of overall score and locomotor and ball skills subscales' scores are presented in Table 2. Significant differences in the test factor (live vs. slow-motion) were found in locomotor subscale in repeated measures analysis (live: 33.7 ± 4.0; slow-motion: 33.2 ± 3.9; $p = 0.010$). The post-hoc Bonferroni analysis revealed significant differences between live and slow-motion evaluations in both NR (Rater A: $p = 0.014$/Rater B: $p = 0.036$) in the locomotor subscale. Same differences were found in the ball subscale in one of the NR (Rater B: $p = 0.001$) and overall score in the same NR (Rater B: $p < 0.001$) and one ER (Rater D: $p = 0.038$).

3.2. Inter-Rater Reliability: Novice vs. Experts

In general, poor-to-moderate inter-rater reliability was found in both locomotor and ball skills, with null changes comparing live vs. slow-motion between each pair of raters. Tables 3 and 4 present inter-rater ICC results for each skill. Figure 1 presents the ICC associations with locomotor and ball subscales and overall scores.

Regarding locomotor skills, reliability were moderate-to-good; for most cases, independent of the rater's experience or background. Gallop was the skill with the best inter-rater reliability; horizontal jump and skip reached ICC values over 0.5 in most comparisons. Hop was the locomotor skill with lower inter-rater reliability.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 1. Intra-rater reliability *: live vs. slow-motion: Subscales, skills and overall scores.

	NR		NR-PE	ER		Good-to-Excellent or Excellent
	Rater A	**Rater B**	**Rater C**	**Rater D**	**Rater E**	
Locomotor skills						**Number of raters**
Run	0.304 (−0.409–0.676)	0.660 (0.236–0.850)	0.900 (0.773–0.956)	**0.913 (0.805–0.961)**	0.942 (0.869–0.975)	3/5
Slide	0.808 (0.567–0.915)	0.921 (0.822–0.965)	**0.966 (0.924–0.985)**	**----- †**	0.714 (0.365–0.873)	2/4
Gallop	0.950 (0.888–0.978)	0.925 (0.823–0.968)	**0.966 (0.923–0.985)**	**0.986 (0.969–0.994)**	**0.976 (0.944–0.989)**	5/5
Hop	0.825 (0.610–0.922)	0.705 (0.332–0.870)	0.936 (0.855–0.972)	**0.990 (0.977–0.996)**	0.884 (0.736–0.949)	2/5
Horizontal jump	0.765 (0.460–0.897)	0.680 (0.166–0.868)	**0.956 (0.900–0.981)**	**0.992 (0.982–0.997)**	0.916 (0.811–0.963)	3/5
Skip	0.892 (0.758–0.952)	0.861 (0.685–0.939)	**1.000**	**0.989 (0.976–0.995)**	**0.958 (0.905–0.981)**	4/5
Locomotor subscale score	0.865 (0.654–0.943)	0.817 (0.592–0.919)	**0.964 (0.918–0.984)**	**0.980 (0.954–0.991)**	0.901 (0.773–0.957)	3/5
Ball skills						
Two-hand strike	**0.959 (0.909–0.982)**	0.905 (0.783–0.958)	0.945 (0.876–0.976)	**0.968 (0.925–0.986)**	0.945 (0.876–0.976)	5/5
One-hand stationary dribble	0.917 (0.813–0.964)	0.871 (0.662–0.947)	0.917 (0.803–0.964)	0.883 (0.739–0.948)	0.910 (0.798–0.960)	3/5
Overhand throw	0.931 (0.844–0.970)	0.887 (0.742–0.950)	0.934 (0.852–0.971)	**0.975 (0.942–0.989)**	0.915 (0.807–0.963)	4/5
Kick	0.659 (0.230–0.849)	0.846 (0.625–0.934)	0.839 (0.537–0.936)	0.948 (0.882–0.977)	**0.982 (0.956–0.992)**	2/5
Forehand strike	0.818 (0.582–0.920)	0.875 (0.720–0.945)	0.891 (0.745–0.953)	0.955 (0.898–0.980)	0.936 (0.856–0.972)	2/5
Two-hand catch	**0.983 (0.961–0.992)**	0.899 (0.770–0.955)	0.908 (0.791–0.959)	**1.000**	**1.000**	5/5
Underhand throw	0.916 (0.809–0.963)	0.698 (0.333–0.865)	**0.966 (0.923–0.985)**	**0.958 (0.901–0.982)**	0.821 (0.597–0.921)	3/5
Ball Skills Subscale score	**0.970 (0.934–0.987)**	0.879 (0.689–0.949)	0.949 (0.884–0.978)	**0.967 (0.915–0.986)**	**0.962 (0.915–0.983)**	4/5
Overall score	0.953 (0.895–0.979)	0.907 (0.716–0.963)	**0.968 (0.927–0.986)**	**0.982 (0.941–0.993)**	**0.961 (0.912–0.983)**	4/5

* ICC: single-measurement, absolute-agreement, 2-way mixed-effects model, 95% confidence interval; **†**: Rater D (ER) coded slide skill with max punctuations in both test (live and slow-motion); In grey, skills with good-to-excellent intra-rater reliability (ICC & lower bound ≥ 0.75); in grey and bold, skills with excellent intra-rater reliability (ICC & lower bound ≥ 0.9).

Table 2. Descriptive statistics of overall score and locomotor and ball skills subscales' scores. Results expressed as mean (standard deviation).

Raters		**Locomotor Subscale**		**Ball Skills Subscale**		**Overall**	
		Live	**S-M**	**Live**	**S-M**	**Live**	**S-M**
NR	A	34.2 (3.4)	33.2 (3.1)	36.0 (5.3)	36.4 (4.9)	70.2 (7.7)	69.5 (6.3)
	B	34.8 (4.3)	33.9 (4.4)	35.2 (4.4)	33.8 (5.0)	70.0 (7.1)	67.6 (7.9)
NR-PE	C	32.5 (4.6)	32.3 (4.5)	33.9 (5.0)	34.8 (4.3)	66.4 (7.9)	67.2 (7.6)
ER	D	32.6 (4.0)	32.9 (4.0)	39.7 (5.6)	40.6 (5.5)	72.2 (8.6)	73.4 (8.4)
	E	34.5 (3.2)	33.8 (3.6)	37.6 (5.0)	37.6 (4.4)	72.2 (6.4)	71.5 (6.3)

NR: Novice raters with no Physical Education background (A and B); NR-PE: Novice rater with Physical Education background (C); ER: Expert raters (D and E).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Int. J. Environ. Res. Public Health **2021**, *18*, 1652

Table 3. Inter-rater reliability * among raters. Locomotor skills.

Raters	Test	Run	Slide †	Gallop	Hop	Horizontal Jump	Skip	ICC ≥ 0.5
Rater AxB	Live	0.074 (−0.323–0.450)	0.739 (0.492–0.876)	**0.768 (0.541–0.891)**	0.525 (0.171–0.758)	0.624 (0.311–0.815)	**0.856 (0.701–0.934)**	5/6
	S-M	−0.002 (−0.390–0.386)	**0.815 (0.624–0.914)**	**0.747 (0.505–0.880)**	0.581 (0.249–0.791)	0.497 (0.135–0.742)	**0.747 (0.506–0.880)**	4/6
Rater AxC	Live	0.106 (−0.294–0.475)	**0.791 (0.581–0.902)**	0.610 (0.290–0.807)	0.477 (0.109–0.730)	0.586 (0.255–0.793)	**0.939 (0.866–0.973)**	4/6
	S-M	0.090 (−0.309–0.462)	0.437 (0.058–0.705)	0.687 (0.407–0.849)	0.251 (−0.152–0.583)	0.688 (0.409–0.849)	**0.805 (0.607–0.909)**	3/6
Rater AxD	Live	0.129 (−0.273–0.493)	-----	**0.801 (0.600–0.907)**	0.496 (0.134–0.742)	0.626 (0.314–0.816)	0.612 (0.294–0.808)	3/5
	S-M	−0.052 (−0.431–0.643)	-----	**0.893 (0.773–0.952)**	0.385 (−0.004–0.672)	0.598 (0.274–0.801)	0.588 (0.259–0.795)	3/5
Rater AxE	Live	−0.022 (0.407–0.369)	0.647 (0.345–0.827)	0.672 (0.383–0.841)	0.297 (−0.103–0.614)	0.364 (−0.028–0.659)	0.710 (0.444–0.861)	2/6
	S-M	−0.106 (−0.474–0.295)	**0.749 (0.509–0.881)**	**0.771 (0.546–0.892)**	0.470 (0.100–0.726)	0.456 (0.082–0.717)	0.735 (0.485–0.874)	3/6
Rater BxC	Live	0.602 (0.279–0.803)	0.722 (0.464–0.867)	**0.883 (0.753–0.947)**	0.393 (0.005–0.678)	0.669 (0.379–0.839)	**0.760 (0.527–0.887)**	5/6
	S-M	0.491 (0.127–0.739)	0.451 (0.076–0.714)	0.723 (0.465–0.868)	0.423 (0.041–0.697)	0.597 (0.272–0.800)	**0.773 (0.550–0.893)**	3/6
Rater BxD	Live	0.471 (0.101–0.726)	-----	**0.874 (0.736–0.943)**	0.563 (0.224–0.781)	0.662 (0.369–0.836)	0.481 (0.114–0.732)	3/5
	S-M	0.629 (0.318–0.817)	-----	**0.814 (0.623–0.914)**	0.447 (0.072–0.712)	0.597 (0.272–0.800)	0.420 (0.038–0.695)	3/5
Rater BxE	Live	0.568 (0.231–0.784)	0.529 (0.177–0.761)	**0.816 (0.626–0.915)**	0.350 (−0.044–0.650)	0.492 (0.129–0.739)	0.732 (0.481–0.872)	3/6
	S-M	0.681 (0.399–0.846)	**0.759 (0.526–0.886)**	**0.808 (0.611–0.910)**	0.692 (0.415–0.851)	0.585 (0.255–0.793)	0.608 (0.288–0.806)	6/6
Rater CxD	Live	0.569 (0.232–0.784)	-----	**0.827 (0.646–0.920)**	0.656 (0.356–0.832)	**0.784 (0.568–0.899)**	0.636 (0.330–0.822)	5/5
	S-M	0.528 (0.176–0.760)	-----	**0.766 (0.538–0.890)**	0.552 (0.209–0.774)	**0.767 (0.539–0.890)**	0.652 (0.353–0.830)	5/5
Rater CxE	Live	0.587 (0.258–0.794)	0.483 (0.117–0.734)	**0.841 (0.672–0.927)**	0.457 (0.084–0.718)	0.407 (0.023–0.687)	0.702 (0.432–0.857)	3/6
	S-M	0.683 (0.401–0.847)	0.640 (0.336–0.824)	0.704 (0.434–0.858)	0.423 (0.042–0.697)	0.591 (0.263–0.796)	0.655 (0.358–0.832)	5/6
Rater DxE	Live	0.551 (0.207–0.774)	-----	**0.766 (0.538–0.890)**	0.183 (−0.221–0.533)	0.637 (0.330–0.822)	0.483 (0.117–0.734)	3/5
	S-M	0.459 (0.086–0.719)	-----	**0.826 (0.645–0.920)**	0.433 (0.054–0.703)	0.658 (0.362–0.833)	0.591 (0.263–0.796)	3/5
All raters	Live	0.423 (0.243–0.629)	0.494 (0.314–0.686)	**0.795 (0.674–0.889)**	0.431 (0.251–0.636)	0.597 (0.425–0.762)	**0.693 (0.540–0.827)**	3/6
	S-M	0.368 (0.192–0.581)	0.499 (0.319–0.690)	**0.770 (0.641–0.875)**	0.468 (0.287–0.666)	0.616 (0.446–0.775)	0.652 (0.489–0.800)	3/6

Rater A and B: Novice raters with no Physical Education background; Rater C: Novice rater with Physical Education background; Rater D & E: Expert raters. * ICC: single-measurement, consistency, 2-way random-effects model, 95% confidence interval; † Rater D (ER) coded slide skill with max punctuations in both test (live and slow-motion); In grey, skills with ICC ≥ 0.5; in grey and bold, skills with at least moderate inter-rater reliability (ICC & lower bound ≥ 0.5).

Table 4. Inter-rater reliability * among raters. Ball skills.

Raters	Test	Two-Hand Strike	One-Hand Stationary Dribble	Overhand Throw	Kick	Forehand Strike	Two-Hand Catch	Underhand Throw	ICC ≥ 0.5
Rater AxB	Live	**0.779 (0.561–0.896)**	0.372 (−0.019–0.664)	0.548 (0.204–0.772)	0.393 (0.006–0.678)	0.632 (0.323–0.819)	0.669 (0.379–0.839)	0.348 (−0.046–0.649)	4/7
	S-M	**0.751 (0.512–0.882)**	**0.778 (0.559–0.896)**	0.695 (0.420–0.853)	0.289 (−0.112–0.609)	0.639 (0.333–0.823)	0.621 (0.306–0.813)	0.598 (0.272–0.800)	6/7
Rater AxC	Live	**0.805 (0.607–0.909)**	0.396 (0.009–0.680)	**0.783 (0.567–0.898)**	0.579 (0.246–0.790)	0.569 (0.232–0.784)	**0.876 (0.739–0.943)**	0.589 (0.261–0.796)	6/7
	S-M	**0.788 (0.576–0.901)**	0.631 (0.322–0.819)	0.660 (0.365–0.834)	0.478 (0.110–0.730)	0.674 (0.387–0.842)	**0.929 (0.845–0.968)**	0.475 (0.106–0.729)	5/7
Rater AxD	Live	**0.754 (0.517–0.883)**	0.736 (0.488–0.875)	0.586 (0.255–0.793)	0.367 (−0.025–0.661)	0.522 (0.167–0.757)	0.392 (0.005–0.677)	0.594 (0.267–0.798)	5/7
	S-M	0.630 (0.320–0.818)	0.698 (0.426–0.855)	0.445 (0.069–0.710)	0.532 (0.181–0.763)	0.549 (0.204–0.772)	0.389 (0.001–0.675)	0.526 (0.174–0.759)	5/7
Rater AxE	Live	0.660 (0.366–0.835)	0.560 (0.219–0.779)	0.570 (0.233–0.785)	0.500 (0.139–0.744)	0.080 (−0.318–0.454)	0.308 (−0.091–0.622)	0.264 (−0.138–0.592)	4/7
	S-M	0.732 (0.480–0.872)	0.545 (0.199–0.770)	0.476 (0.107–0.729)	0.412 (0.028–0.690)	0.378 (−0.012–0.668)	0.297 (−0.103–0.614)	0.322 (−0.076–0.631)	3/7
Rater BxC	Live	0.698 (0.426–0.855)	**0.900 (0.787–0.955)**	**0.749 (0.508–0.881)**	0.404 (0.019–0.685)	0.647 (0.346–0.828)	0.600 (0.276–0.801)	0.469 (0.099–0.725)	5/7
	S-M	0.681 (0.398–0.846)	0.735 (0.486–0.874)	**0.776 (0.554–0.895)**	0.462 (0.090–0.721)	0.470 (0.100–0.726)	0.705 (0.436–0.858)	0.614 (0.297–0.810)	5/7
Rater BxD	Live	0.544 (0.198–0.770)	0.337 (−0.058–0.642)	0.462 (0.090–0.721)	0.740 (0.493–0.876)	0.691 (0.414–0.851)	0.308 (−0.091–0.622)	0.540 (0.192–0.767)	4/7
	S-M	0.582 (0.250–0.791)	0.643 (0.340–0.825)	0.373 (−0.018–0.665)	**0.778 (0.559–0.896)**	0.611 (0.292–0.808)	0.228 (−0.176–0.566)	0.636 (0.330–0.822)	5/7

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4. *Cont.*

Raters	Test	Two-Hand Strike	One-Hand Stationary Dribble	Overhand Throw	Kick	Forehand Strike	Two-Hand Catch	Underhand Throw	ICC ≥ 0.5
Rater BxE	Live	0.652 (0.354–0.830)	0.385 (−0.004–0.673)	0.596 (0.271–0.800)	0.533 (0.182–0.763)	0.361 (−0.032–0.657)	0.302 (−0.097–0.618)	0.292 (−0.109–0.611)	3/7
	S-M	**0.809 (0.614–0.911)**	0.580 (0.248–0.790)	0.608 (0.288–0.806)	0.534 (0.184–0.764)	0.482 (0.116–0.733)	0.227 (−0.177–0.565)	0.285 (−0.117–0.606)	4/7
Rater CxD	Live	0.716 (0.453–0.864)	0.426 (0.045–0.699)	0.610 (0.290–0.807)	0.389 (0.001–0.675)	0.600 (0.276–0.802)	0.410 (0.026–0.689)	0.711 (0.447–0.892)	4/7
	S-M	0.487 (0.121–0.736)	0.593 (0.267–0.798)	0.684 (0.402–0.847)	0.339 (−0.057–0.642)	0.565 (0.226–0.782)	0.313 (−0.086–0.625)	0.665 (0.374–0.837)	4/7
Rater CxE	Live	0.682 (0.400–0.846)	0.491 (0.127–0.739)	0.589 (0.261–0.796)	0.412 (0.028–0.690)	0.357 (−0.036–0.655)	0.303 (−0.097–0.618)	0.248 (−0.155–0.580)	2/7
	S-M	0.615 (0.298–0.810)	0.477 (0.109–0.730)	0.570 (0.234–0.785)	0.378 (−0.012–0.668)	0.409 (0.024–0.688)	0.222 (−0.181–0.562)	0.377 (−0.014–0.667)	2/7
Rater DxE	Live	0.574 (0.239–0.787)	0.581 (0.249–0.791)	0.505 (0.146–0.747)	0.600 (0.277–0.802)	0.241 (−0.162–0.575)	0.586 (0.255–0.793)	0.176 (−0.228–0.528)	5/7
	S-M	0.530 (0.187–0.762)	0.525 (0.172–0.759)	0.343 (−0.052–0.645)	0.615 (0.297–0.810)	0.287 (−0.114–0.607)	0.586 (0.255–0.793)	0.466 (0.095–0.724)	3/7
All raters	Live	**0.686 (0.531–0.823)**	0.518 (0.339–0.705)	0.595 (0.423–0.761)	0.500 (0.320–0.691)	0.470 (0.289–0.667)	0.496 (0.316–0.688)	0.452 (0.271–0.653)	4/7
	S-M	0.652 (0.489–0.800)	0.627 (0.459–0.783)	0.558 (0.381–0.734)	0.498 (0.318–0.689)	0.502 (0.322–0.692)	0.496 (0.316–0.688)	0.506 (0.326–0.695)	5/7

Rater A and B: Novice raters; Rater C: Novice rater with Physical Education background; Rater D & E: Expert raters. * ICC: single-measurement, consistency, 2-way random-effects model, 95% confidence interval; In grey, skills with ICC ≥ 0.5; in grey and bold, skills with at least moderate inter-rater reliability (ICC & lower bound ≥ 0.5).

Regarding ball skills, the two-hand strike was the skill with higher inter-rater reliability, ranging between poor-to-good and moderate-to-excellent. The two-hand catch was the skill with the lower inter-rater reliability.



Figure 1. *Cont.*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

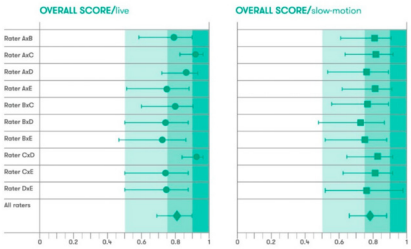


Figure 1. Inter-rater intraclass correlation coefficients (ICC) values of locomotor and ball skills subscales and overall score.

Ball skills subscale and overall scores are slightly higher in ER compared with NR and NR-PE (Table 2). The repeated measures analysis showed no differences in the interactions for both viewing modes and raters factors for the locomotor subscale. However, regarding ball skills, one of the experts, rater (Rater D), coded significantly higher in live evaluation than one NR (Rater B: $p = 0.024$) and NR-PE ($p = 0.001$). Same differences were found in slow-motion evaluation between Rater D and both NR (Rater A: $p = 0.029$/Rater B: $p = < 0.001$) and NR-PE ($p = 0.001$). Finally, overall scores registered only significant differences between one ER (Rater D) and NR-PE in slow-motion evaluation ($p = 0.031$).

4. Discussion

The purpose of this study was to analyze the intra- (live and slow-motion) and inter-rater [expert and novice (one of them with physical education background)] of TGMD-3. Briefly, locomotor and ball skills subscales and overall scores showed moderate-to-excellent intra-rater reliability. The inter-rater reliability showed moderate-to-good scores for the ball skills subscale and overall scores for both live and slow-motion and poor-to-good reliability for the locomotor subscale also for both assessing modes.

The overall score for intra-rater reliability was, at least, moderate-to-excellent. However, ER and NR-PE reached higher levels of reliability than NR, with excellent intra-rater ICC values. In this sense, previous studies also showed good-to-excellent or excellent intra-rater reliability in the TGMD-2 [24–27] and the TGMD-3 [28–32]. For locomotor and ball skills subscales' scores, although intra-rater reliability of NR was at least moderate-to-excellent, ER and NR-PE achieved no less than good-to-excellent. High agreement was also reported in both subscales in the TGMD-2 [24–27] and the TGMD-3 [28,30–33].

In the stringent analysis performed by skill, all the raters achieved good-to-excellent or excellent intra-rater reliability for gallop, two-hand strike and two-hand catch. Nevertheless, it was observed more differences between raters in intra-rater the individual skills, ICC values compared to overall score or subscales scores. Previous studies also found variations in the intra-rater reliability agreement across raters [31], especially when skills are examined individually [10], obtaining results ranging from poor-to-excellent. In the present study, intra-rater reliability varied from moderate-to-excellent to excellent for ER and NR-PE in all skills (except Slide for Rater E). Variability in ICC values was higher in NR, with moderate-to-good or moderate-to-excellent intra-rater reliability in half of the skills.

Regarding inter-rater reliability, our results show that the degree of consistency varies considerably between raters on the TGMD-3 skills. As observed, the variations in inter-rater reliability were more visible in the analysis by skill. The ICC's results for the overall test, locomotor and ball skills subscales' scores were over 0.5 in almost all comparisons. The inter-rater reliability for locomotor skills, among five raters, was poor-to-moderate for run,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

slide and hop (both live and slow-motion); and poor-to-good for the skip in slow-motion and the horizontal jump (both live and slow-motion); and moderate-to-good for the skip in live viewing and the gallop in both viewing modes. Overall, the skill with the higher reliability among raters was the gallop achieving an ICC > 0.5 in all comparisons, reaching moderate reliability in 15/20 comparisons between raters. In contrast, the lower results of reliability were found for the hop and run skill. Kim et al. [34] have also reported the gallop as the skill with higher reliability results and the run skill as the one with the lowest reliability coefficients. In contrast, Maeng et al. [31] and Ayán et al. [35] achieved the most robust reliability in the hop and the weakest in the gallop, contrary to our findings.

Regarding the ball skills, the inter-rater reliability was poor-to-moderate for one-hand stationary dribble (live), overhand throw (slow-motion), kick, forehand strike, two-hand catch and underhand throw (both live and-slow motion); poor-to-good for two-hand strike (slow-motion), one-hand stationary dribble (slow-motion) and overhand throw (live); and moderate-to-good for the two-hand strike in live viewing. In general, the two-hand strike was the skill with the best reliability among raters reaching an ICC > 0.5 and moderate reliability in 19/20 comparisons. The lowest reliability coefficients achieved were in two-hand catch skill. This result is in concordance with previous studies [36] but again, contrary to others that found the two-hand catch as the skill with stronger inter-rater reliability [10].

Our results show a lower agreement on the locomotor skills compared to the ball skills. Rintala et al. [10] and Palmer and Brian [22], who reported similar results in TGMD-2, suggested that before using TGMD, further training with the locomotor skills may be needed, our results support this contention. On the other hand, it seems that the raters' background and experience did not influence the inter-rater reliability, taking into account that inter-rater reliability varied not only among NR and ER raters but also between NR each other and ER. Contrary, Palmer and Brian [22] found that novice raters could not achieve a significant agreement with expert raters assessing the TGMD-2 test. However, it has been shown that novice raters reached a significantly similar agreement to experts assessing measure movements with the Landing Error Scoring System [37] and Functional Movement Screen [38].

The variation found in inter-rater reliability in our study could be due to different interpretations of each skill evaluation criteria. Some of them give rise to subjectivity and vary according to the concrete interpretation of the raters. For example, in the hop skill, the second criterion, "Foot of non-hopping leg remains behind hopping leg" might be interpreted differently; it might consider scoring 1 if non-hopping foot always remained behind hopping leg. Nevertheless, it might be considered scoring 1 if the child, although not always, executed the criterion correctly most of the time. Hence, it becomes necessary to obtain pre-assessment agreements between raters or professionals who pretend to evaluate FMS in children. It has been suggested that more training may be necessary for those skills more difficult to assess and even that some of the criteria might be re-defined to make them more objective [10,20,31,34,36]; our results provide further support for this recommendation.

Besides, a recent systematic review about TGMD-2 and 3 reliability showed at least good-to-excellent ICC values for inter-rater reliability in both subscales and the overall score, higher than those obtained in this study. However, these differences might be regarding the methodology adopted or the statistical analysis used [20]. In the case of ICC, used in many studies, several aspects make it challenging to compare the results. First, investigators have to choose an ICC model (One-Way Random-Effects Model, Two-Way Random-Effects Model or Two-Way Mixed-Effects Model), ICC definition (absolute agreement or consistency) and ICC type (single or mean measurement) according to the study. Second, there are no standards values to report reliability. Third, some studies did not report ICC model/type/definition or reported reliability based on ICC values with no mention of the confidence interval. Our study followed the flowchart proposed by Koo et al. [23] for selecting model, type and definition. Many studies have used different

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

models, in most cases, less demanding, which might explain the higher reliability results obtained compared to our study.

About to ICC values and 95% confidence intervals, interpretation was also based on Koo et al. [23] recommendations. In this respect, ICC values do not describe reliability in itself but it is necessary to express reliability in terms of lower and upper bounds of the confidence interval. Finally, we considered excellent reliability if an ICC equal to or higher than 0.9 was obtained, while other ratings considered it above 0.75 [39,40]. All this makes it difficult to compare the reliability used in different studies since this variability causes an interpretation problem.

Limitations

This study is not free of limitations. Our reliability study was carried out assessing 25 healthy children. The sample was small, so the results should be interpreted with caution. Besides, we tried to simulate the live assessment by displaying the digital records only once and playing the video at a normal speed. This procedure does not represent precisely the reality of a live assessment by professionals who have to score the children's performance. However, from a research point of view, it is a useful way for different evaluators to carry out the evaluation at any time.

5. Conclusions

The TGMD-3 battery showed moderate-to-excellent intra-rater reliability for overall score, locomotor and ball skills subscales' scores and moderate-to-good inter-rater reliability for overall and ball skills subscale scores and poor-to-good for locomotor subscale. The expert raters and the novice rater with physical education background achieved stronger intra-rater reliability than novice raters and inter-rater reliability did not seem to be influenced by the raters' background and experience. The viewing modes also seemed not to influence on reliability but further investigation is needed in this regard, in the sight of the study limitations. Higher variability in both intra- and inter-rater reliability was found when analyzing the skills separately.

In conclusion, our high results in terms of intra-rater reliability but lower in the case of inter-rater, suggest that each skill's criteria can be interpreted differently. For clinical practice, it would be recommended that raters reach an agreement before the assessment to avoid subjective interpretations that might distort the results. In addition, a revision of some criteria might be needed to let research replication.

Author Contributions: Conceptualization, E.R., C.V.-C. and C.A.-G.; methodology, A.C.-F., E.R., N.C.V. and C.A.-G.; validation, C.V.-C. and C.A.-G.; formal analysis, R.B.-F. and C.A.-G.; investigation, A.C.-F. and J.E.R.-F.; resources, E.R., J.R.-D. and R.B.-F.; data curation, A.C.-F., J.R.-D. and C.A.-G.; writing—original draft preparation, A.C.-F. and C.A.-G.; writing—review and editing, A.C.-F., E.R., N.C.V. and C.A.-G.; visualization, C.V.-C.; supervision, C.A.-G.; project administration, E.R., C.V.-C. and C.A.-G.; funding acquisition, A.C.-F., E.R., C.V.-C. and C.A.-G. All authors have read and agreed to the published version of the manuscript.

Funding: The ALFA-MOV Project is funded by FEDER/Spanish Ministry of Science, Innovation and Universities—National Agency of Investigation/Project RTI2018-096106-A-I00. AC-F (FPU19/02017) is recipient of a predoctoral fellowship by the Spanish Ministry of Science, Innovation and Universities.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the University of Vigo, Spain (protocol code: 17-0320; date of approval: 11 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We would like to thank the participants who made this study possible, as well as parents/guardians and schoolteachers.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gallahue, D.L.; Donnelly, F.C. *Developmental Physical Education for All Children*, 4th ed.; Human Kinetics: Champaing, IL, USA, 2007.
2. Gallahue, D.; Ozmun, J.; Goodway, J. *Understanding Motor Development: Infants, Children, Adolescents, Adults*, 7th ed.; McGraw-Hill: New York, NY, USA, 2012.
3. Logan, S.W.; Ross, S.M.; Chee, K.; Stodden, D.F.; Robinson, L.E. Fundamental motor skills: A systematic review of terminology. *J. Sports Sci.* **2018**, *36*, 781–796. [CrossRef]
4. Lubans, D.R.; Morgan, P.J.; Cliff, D.P.; Barnett, L.M.; Okely, A.D. Fundamental Movement Skills in Children and Adolescents. Review of associated health benefits. *Sports Med.* **2010**, *40*, 1019–1035. [CrossRef]
5. Cattuzzo, M.T.; Henrique, R.D.S.; Ré, A.H.N.; De Oliveira, I.S.; Melo, B.M.; Moura, M.D.S.; De Araújo, R.C.; Stodden, D. Motor competence and health related physical fitness in youth: A systematic review. *J. Sci. Med. Sport* **2016**, *19*, 123–129. [CrossRef]
6. Barnett, L.M.; Lai, S.K.; Veldman, S.L.C.; Hardy, L.L.; Cliff, D.P.; Morgan, P.J.; Zask, A.; Lubans, D.R.; Shultz, S.P.; Ridgers, N.D.; et al. Correlates of Gross Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Med.* **2016**, *46*, 1663–1688. [CrossRef]
7. Robinson, L.E.; Stodden, D.F.; Barnett, L.M.; Lopes, V.P.; Logan, S.W.; Rodrigues, L.P.; D'Hondt, E. Motor Competence and its Effect on Positive Developmental Trajectories of Health. *Sports Med.* **2015**, *45*, 1273–1284. [CrossRef] [PubMed]
8. Haapala, E.A. Cardiorespiratory Fitness and Motor Skills in Relation to Cognition and Academic Performance in Children—A Review. *J. Hum. Kinet.* **2013**, *36*, 55–68. [CrossRef] [PubMed]
9. Costa, H.J.T.; Barcala-Furelos, R.; Abelairas-Gómez, C.; Arufe-Giráldez, V. The Influence of a Structured Physical Education Plan on Preschool Children's Psychomotor Development Profiles. *Australas. J. Early Child.* **2015**, *40*, 68–77. [CrossRef]
10. Rintala, P.; Sääkslahti, A.; Iivonen, S. Reliability Assessment of Scores from Video-Recorded TGMD-3 Performances. *J. Mot. Learn. Dev.* **2017**, *5*, 59–68. [CrossRef]
11. Logan, S.W.; Robinson, L.E.; Rudisill, M.E.; Wadsworth, D.D.; Morera, M. The comparison of school-age children's performance on two motor assessments: The Test of Gross Motor Development and the Movement Assessment Battery for Children. *Phys. Educ. Sport Pedagog.* **2014**, *19*, 48–59. [CrossRef]
12. Piek, J.P.; Dawson, L.; Smith, L.M.; Gasson, N. The role of early fine and gross motor development on later motor and cognitive ability. *Hum. Mov. Sci.* **2008**, *27*, 668–681. [CrossRef] [PubMed]
13. Skinner, R.A.; Piek, J.P. Psychosocial implications of poor motor coordination in children and adolescents. *Hum. Mov. Sci.* **2001**, *20*, 73–94. [CrossRef]
14. Burton, A.; Miller, D. *Movement Skill Assessment*; Human Kinetics: Champaing, IL, USA, 1998.
15. Yun, J.; Shapiro, D.R. A Quantitative Approach to Movement Skill Assessment for Children with Mental Retardation. *Adapt. Phys. Act. Q.* **2004**, *21*, 269–280. [CrossRef]
16. Clark, J.E. Motor Development. In *Encyclopedia of Human Behavior*, 3rd ed.; Ramachandran, V.S., Ed.; Academic Press: New York, NY, USA, 1994; pp. 245–255.
17. Payne, V.G.; Isaacs, L.D. *Human Motor Development*; McGraw-Hill: New York, NY, USA, 2008.
18. Ulrich, D.A. *Test of Gross Motor Development*, 2nd ed.; PRO-ED: Austin, TX, USA, 2000.
19. Ulrich, D.A. *Test of Gross Motor Development*, 3rd ed.; PRO-ED: Austin, TX, USA, 2019.
20. Rey, E.; Carballo-Fazanes, A.; Varela-Casal, C.; Abelairas-Gómez, C. Reliability of the test of gross motor development: A systematic review. *PLoS ONE* **2020**, *15*, e0236070. [CrossRef]
21. Akuffo, P.B.; Hodge, S.R. Roles and Responsibilities of Adapted Physical Education Teachers in an Urban School District. *Educ. Urban Soc.* **2008**, *40*, 243–268. [CrossRef]
22. Palmer, K.K.; Brian, A. Test of Gross Motor Development-2 Scores Differ Between Expert and Novice Coders. *J. Mot. Learn. Dev.* **2016**, *4*, 142–151. [CrossRef]
23. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]
24. Aye, T.; Oo, K.S.; Khin, M.T.; Kuramoto-Ahuja, T.; Maruyama, H. Reliability of the test of gross motor development second edition (TGMD-2) for Kindergarten children in Myanmar. *J. Phys. Ther. Sci.* **2017**, *29*, 1726–1731. [CrossRef]
25. Capio, C.M.; Eguia, K.F.; Simons, J. Test of gross motor development-2 for Filipino children with intellectual disability: Validity and reliability. *J. Sports Sci.* **2015**, *34*, 10–17. [CrossRef]
26. Farrokhi, A.; Zareh, Z.; Karimi, A.; Kazemnejad, A.; Ilbeigi, S. Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3–10 aged children of Tehran City. *J. Phys. Educ. Sport Manag.* **2014**, *5*, 18–28. [CrossRef]
27. Houwen, S.; Hartman, E.; Jonker, L.; Visscher, C. Reliability and Validity of the TGMD-2 in Primary-School-Age Children with Visual Impairments. *Adapt. Phys. Act. Q.* **2010**, *27*, 143–159. [CrossRef]
28. Allen, K.A.; Bredero, B.; Van Damme, T.; Ulrich, D.A.; Simons, J. Test of Gross Motor Development-3 (TGMD-3) with the Use of Visual Supports for Children with Autism Spectrum Disorder: Validity and Reliability. *J. Autism Dev. Disord.* **2017**, *47*, 813–833. [CrossRef]
29. Estevan, I.; Molina-García, J.; Queralt, A.; Álvarez, O.; Castillo, I.; Barnett, L. Validity and Reliability of the Spanish Version of the Test of Gross Motor Development–3. *J. Mot. Learn. Dev.* **2017**, *5*, 69–81. [CrossRef]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

30. Mohammadi, F.; Bahram, A.; Khalaji, H.; Ulrich, D.; Ghadiri, F. Evaluation of the Psychometric Properties of the Persian Version of the Test of Gross Motor Development–3rd Edition. *J. Mot. Learn. Dev.* **2019**, *7*, 106–121. [CrossRef]
31. Maeng, H.; Webster, E.K.; Pitchford, E.; Ulrich, D. Inter- and Intrarater Reliabilities of the Test of Gross Motor Development—Third Edition among Experienced TGMD-2 Raters. *Adapt. Phys. Act. Q.* **2017**, *34*, 442–455. [CrossRef]
32. Valentini, N.C.; Zanella, L.W.; Webster, E.K. Test of Gross Motor Development—Third Edition: Establishing Content and Construct Validity for Brazilian Children. *J. Mot. Learn. Dev.* **2017**, *5*, 15–28. [CrossRef]
33. Wagner, M.O.; Webster, E.K.; Ulrich, D.A. Psychometric Properties of the Test of Gross Motor Development, Third Edition (German Translation): Results of a Pilot Study. *J. Mot. Learn. Dev.* **2017**, *5*, 29–44. [CrossRef]
34. Kim, Y.; Park, I.; Kang, M. Examining Rater Effects of the TGMD-2 on Children with Intellectual Disability. *Adapt. Phys. Act. Q.* **2012**, *29*, 346–365. [CrossRef]
35. Ayán, C.; Cancela, J.; Sanchez-Lastra, M.A.; Carballo-Roales, A.; Domínguez-Meis, F.; Redondo-Gutiérrez, L. Fiabilidad y Validez de la Batería TGMD-2 en Población Española. *Rev. Iberoam. Diagnóstico Eval.* **2019**, *50*, 21–34. [CrossRef]
36. Barnett, L.M.; Minto, C.; Lander, N.; Hardy, L.L. Interrater reliability assessment using the Test of Gross Motor Development-2. *J. Sci. Med. Sport* **2014**, *17*, 667–670. [CrossRef] [PubMed]
37. Onate, J.; Cortes, N.; Welch, C.; Van Lunen, B.L. Expert Versus Novice Interrater Reliability and Criterion Validity of the Landing Error Scoring System. *J. Sport Rehabil.* **2010**, *19*, 41–56. [CrossRef]
38. Minick, K.I.; Kiesel, K.B.; Burton, L.; Taylor, A.; Plisky, P.; Butler, R.J. Interrater Reliability of the Functional Movement Screen. *J. Strength Cond. Res.* **2010**, *24*, 479–486. [CrossRef] [PubMed]
39. Fleiss, J.L. *Design and Analysis of Clinical Experiments*; John Wiley & Sons: New York, NY, USA, 1986.
40. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [CrossRef]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ANEXO 2. ESTANCIAS

Instituto Politécnico de Viana do Castelo
Escola Superior
de Desporto e Lazer

17 June 2022

Dear Sir/Madam,

CERTIFICATE OF RESEARCH STAY: AIDA CARBALLO FANZANES

This is a letter to certificate the visit and contribution of Aida Carballo Fanzanes to the Escola Superior de Desporto e Lazer de Melgaço, Instituto Politécnico de Viana do Castelo, Portugal between 15 March and 15 June 2022.

Aida joined us as a self-funded temporary visiting researcher. During her work program for this visit Aida was involved on several ongoing projects, namely: in the organization and collection of annual data of the Melgaço Youth Growth Study; the development and validation of the Motor Competence Assessment (MCA); the collection of pilot data on the PACETECH prototype; and the preparation of the joint study with the Universidad de Vigo on virtual motor competence. Furthermore, Aida Carballo participated in all aspects of the academic life of the ESDL, being recognized by everybody as a valuable member of the ESDL community.

During her stay, Aida Carballo submitted and was accepted for presenting her work in the 17º Seminário em Desenvolvimento Motor da Criança, Faro, Portugal; started the writing of a scientific paper relative to the data collected in the Melgaço Youth Growth Study; presented her research in the Sports Training Master course; participated in the data collection of the Movement ABC that took place in Campo Maior, Portugal; and was involved in the organization of the 1st International Sports Meeting of the ESDL.

Aida Carballo work, participation, and opinion was deeply appreciated by everybody that contacted her during her stay in the ESDL.

Personally, I feel grateful for having the chance to work with Aida during her stay, and hopefully in times to come. Aida's happiness and wiliness in all tasks, allied to her strong scientific preparation and will to know, certainly indicate a bright future in the academia.

Yours faithfully

Assinado por: **Luís Paulo Lopes Brandão Areosa Rodrigues**
Num. de Identificação: 05919349
Data: 2022.06.17 13:12:54+01'00'

Professor Luis Paulo Rodrigues
Professor Coordenador com Agregação

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

VNIVERSITAT DE VALÈNCIA Facultat de Magisteri
Departament de Didàctica
de l'Expressió Musical, Plàstica i Corporal

CERTIFICADO DE ESTANCIA DE INVESTIGACIÓN

CERTIFICADO DE ESTANCIA DEL CENTRO RECEPTOR

APELLIDOS, NOMBRE: **Carballo Fazanes, Aida**

A CUMPLIMENTAR POR EL RESPONSABLE DEL CENTRO DE TRABAJO

EL ABAJO FIRMANTE CERTIFICA QUE LA INVESTIGADORA A LA QUE SE REFIERE ESTE DOCUMENTO HA DESARROLLADO SU TRABAJO DE INVESTIGACIÓN EN EL CENTRO DE TRABAJO DURANTE EL PERIODO SEÑALADO A CONTINUACIÓN

CENTRO RECEPTOR: **Departamento de Didáctica de la Expresión Musical, Plástica y Corporal. Universidad de Valencia.**

NOMBRE DEL INVESTIGADOR RESPONSABLE: **Dr. Javier Molina García. Grupo de Investigación AFIPS (GIUV2013-016)**

NOMBRE DEL PROFESOR RESPONSABLE DEL DEPARTAMENTO: **Dr. Daniel Martos García. Director del Departamento de Didáctica de la Expresión Musical, Plástica y Corporal.**

PERIODO DE ESTANCIA REALIZADO:
DESDE: **01-febrero-2022** HASTA: **01-marzo-2022**

Valencia, 1-marzo-2022

MOLINA GARCIA JAVIER - 48480013S
Firmado digitalmente por MOLINA GARCIA JAVIER - 48480013S
Fecha: 2022.03.01 07:07:31 +01'00'

Fdo: Javier Molina García

MARTOS GARCIA DANIEL - 53095181A
Firmado digitalmente por MARTOS GARCIA DANIEL - 53095181A
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-53095181A, givenName=DANIEL, sn=MARTOS GARCIA, cn=MARTOS GARCIA DANIEL - 53095181A
Fecha: 2022.03.01 07:24:21 +01'00'

Daniel Martos García

(Firma del investigador responsable)

Apartat de Correus 22.045
46071 València (Espanya)
Tf: 96 386 44 84 Fax :96 398 32 10
http://www.uv.es/didexpr
dep.didactica.expressio@uv.es

Avda. dels Tarongers, 4
46022 València

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

VNIVERSITAT D VALÈNCIA Facultat de Magisteri
Departament de Didàctica
de l'Expressió Musical, Plàstica i Corporal

CERTIFICADO DE ESTANCIA DE INVESTIGACIÓN

CERTIFICADO DE ESTANCIA DEL CENTRO RECEPTOR

APELLIDOS, NOMBRE: **Carballo Fazanes, Aida**

A CUMPLIMENTAR POR EL RESPONSABLE DEL CENTRO DE TRABAJO

EL ABAJO FIRMANTE CERTIFICA QUE EL INVESTIGADOR AL QUE SE REFIERE ESTE DOCUMENTO HA DESARROLLADO SU TRABAJO DE INVESTIGACIÓN EN EL CENTRO DE TRABAJO DURANTE EL PERIODO SEÑALADO A CONTINUACIÓN

CENTRO RECEPTOR: **Departamento de Didáctica de la Expresión Musical, Plástica y Corporal. Universidad de Valencia**

NOMBRE DEL INVESTIGADOR RESPONSABLE: **Profesor Titular de Universidad Dr. Daniel Martos García**

CARGO: **Director del Departamento de Didáctica de la Expresión Musical, Plástica y Corporal**

PERIODO DE ESTANCIA REALIZADO:

DESDE:12-abril-2021 HASTA: 13-junio-2021

Valencia, 11-junio-2021

Fdo: Daniel Martos García

(Stamp/Signature of the host professor in charge)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ANEXO 3. Comité de Ética

Universidade Vigo

Facultad de CC de la Educación y del Deporte
Pontevedra.

Yo, Carlos Ayán Pérez en calidad de Secretario del Comité de Ética de la Facultad de Ciencias de la Educación y del Deporte,

HAGO CONSTAR:

Que los siguientes proyectos de investigación fueron aprobados por dicho Comité de Ética tras previa revisión de los mismos:

- **Proyecto:** Relaciones entre competencia motriz, competencia motriz percibida y actividad física en niños y niñas de educación infantil y primaria.
  Investigador Principal: Ezequiel Rey Eiras
  Código:17-0320
- **Proyecto:** Competencia motora real y percibida. Estudio en adolescentes
  Investigador Principal: Ezequiel Rey Eiras
  Código:09-0721

Carlos Ayán Pérez

Firmado digitalmente por Carlos Ayán Pérez
Fecha: 2022.09.07 22:37:17 +02'00'

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



El desarrollo de las habilidades motrices básicas (HMB) es una fase crucial del desarrollo integral infantil con influencia en la adquisición de un estilo de vida saludable. Sin embargo, la evaluación de las HMB es una tarea pendiente. Esta tesis aporta resultados que podrían servir de base para incrementar y mejorar esta evaluación. Se presenta una revisión sistemática sobre la fiabilidad del Test of Gross Motor Development. Además, se aportan resultados novedosos de fiabilidad entre evaluadores de diferentes perfiles, con varios métodos de evaluación, y tras el acuerdo de los evaluadores sobre los criterios de rendimiento de las HMB previo a la evaluación. En base a estos hallazgos, se ha desarrollado Alfamov, aplicación que permite una evaluación sencilla, fiable y sistematizada de las HMB en la infancia.