
Tesis doctoral

*Cuestionarios de actividad física en edad preescolar
y escolar. Estudio de diseño, validez y fiabilidad del
Physical Activity Questionnaire – Young Children
(PAQ-YC)*

Marta Amor Barbosa



Aquesta tesi doctoral està subjecta a la licència [Reconeixement-NoComercial-SenseObraDerivada 4.0 Internacional \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Esta tesis doctoral está sujeta a la licencia [Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

This doctoral thesis is licensed under the [Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Cuestionarios de actividad física en edad preescolar
y escolar. Estudio de diseño, validez y fiabilidad del
Physical Activity Questionnaire – Young Children
(PAQ-YC)

Tesis doctoral. Programa de Doctorado en Ciencias de la Salud

Autor/a:

Marta Amor Barbosa

Directores:

Dra. Caritat Bagur Calafat

Dr. Ferran Rosés i Noguer

Universitat Internacional de Catalunya

Facultad de Medicina y Ciencias de la Salud

Departamento de Fisioterapia

Barcelona, 2021

“To measure is to know. If you can not
measure it, you can not improve it.”

- Lord Kelvin -

Agradecimientos

Quisiera empezar agradeciendo el esfuerzo y el apoyo de todas las personas que han sido imprescindibles para el éxito de esta tesis doctoral. Ha sido un camino duro, pero vosotros lo habéis hecho más fácil.

Gracias a mis directores, la Dra. Caritat Bagur y el Dr. Ferran Rosés, por haberme acompañado durante esta etapa, por vuestro rigor y vuestra dedicación.

Gracias a la Dra. Montserrat Girabent, por haber invertido su tiempo en mí, por su voluntariedad y su paciencia.

Gracias a la Fundación Álex por haber adquirido el material para esta tesis. Especialmente, quiero dar las gracias al Sr. Gabriel Masfurroll, por su forma de entender la inclusión a través del deporte y su fe en el proyecto.

Gracias a los expertos que dedicaron su tiempo para colaborar de forma altruista en el diseño del cuestionario.

Gracias a cada una de las familias que han participado en los estudios, por vuestro esfuerzo desinteresado y vuestro cariño. Gracias en especial a Amaya Zara, Elena Montiel, Ernesto Herrera, Tamara Pano, Luciana Moizé, Mercè Sitjà y a las escuelas CEIP Margalida Florit y CEIP Pere Casanovas, por agilizar y facilitar todo el proceso de recogida de muestra.

Gracias a mis compañeros del departamento de Fisioterapia de la Universitat Internacional de Catalunya, por su apoyo y sus palabras de ánimo, especialmente a la Dra. Almudena Medina, al Sr. Carlos Zárata y a la Sra. Aida Cadellans.

Gracias a mi familia, amigos y pareja, por vuestra paciencia, vuestra comprensión y vuestro apoyo incondicional.

Sin vosotros no hubiera llegado hasta aquí.

A todos, gracias, de corazón.

Resumen

Introducción:

La infancia es un período clave en la consolidación de hábitos saludables, como la práctica de actividad física. Dentro de los instrumentos disponibles para medir este constructo, los cuestionarios ofrecen ciertas ventajas, pero la falta de estudios de alta calidad dificulta la elección del más adecuado. El grupo *International Physical Activity Questionnaire* (IPAQ) cuenta con gran reconocimiento dentro de la comunidad científica y dispone de versiones adaptadas a diferentes poblaciones, países y culturas.

Material y métodos:

Por un lado, se presenta una revisión sistemática sobre propiedades de medición de los cuestionarios IPAQ para preescolares y escolares. Se consultaron las bases de datos MEDLINE, Embase y Sport Discus. Dos revisores evaluaron la calidad metodológica de los estudios utilizando el Checklist de Riesgo de Sesgo de COSMIN. Los resultados se agruparon y calificaron según los criterios para buenas propiedades de medición. La calidad de la evidencia se puntuó utilizando el *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE) modificado.

Por otro lado, se desarrolló el *Physical Activity Questionnaire – Young Children* (PAQ-YC) para niños de 5-7 años. La primera versión fue probada por expertos en un estudio Delphi de 2 rondas. Se estableció como criterio de consenso que el Rango Intercuartil Relativo (RIR) y/o el Coeficiente de Variación (CV) fueran $\leq 20\%$. Se realizaron entrevistas y una prueba de campo en muestras representativas de la población diana. A continuación, se analizó la fiabilidad y el error de medición. El cuestionario se administró en dos ocasiones con un intervalo de 1-3 días. Se presentaron el coeficiente de Kappa o el Índice de Correlación Interclase (ICC) así como el error estándar de medición (SEM) y/o el porcentaje de acuerdo (PoA). Por último, se estudió

la validez de constructo por comparación del PAQ-YC con acelerometría ActiGraph wGT3x-BT (ActiGraph Corp., Pensacola, FL). Los datos brutos se procesaron con el software Actilife como medidas de la magnitud vectorial. Los resultados se presentaron como coeficientes de correlación de Pearson.

Resultados:

Se incluyeron 21 artículos en el análisis cualitativo y 17 artículos en el análisis cuantitativo. Se identificaron los cuestionarios *Early Years - Physical Activity Questionnaire* (EY-PAQ) para menores de 5 años y *Physical Activity Questionnaire - Older Children* (PAQ-C) para niños de 8-14 años. La fiabilidad se calificó como insuficiente para el EY-PAQ (ICC = 0,35-0,47) y como suficiente para el PAQ-C (ICC = 0,79-0,82), ambas con calidad de la evidencia muy baja. La validez de constructo fue insuficiente en todos los casos. Para el EY-PAQ ($r = 0,02-0,30$) la calidad de la evidencia fue muy baja, mientras que para el PAQ-C ($r = 0,30-0,52$) fue moderada. No se pudieron evaluar la validez de contenido, el error de medición y la capacidad de respuesta.

En el diseño del PAQ-YC, las mayores discrepancias en la consulta Delphi ($n = 11-13$) se observaron en el ítem 7 (RIR = 20, CV = 38,73) y en el ítem 8 (RIR = 25, CV = 15,45). Las entrevistas cognitivas ($n = 5$) y la prueba de campo ($n = 104$) confirmaron la versión consensuada por los expertos. La fiabilidad ($n = 83$) mostró índices excelentes (ICC = 0,983-0,991; SEM = 0,28-2,88; PoA = 75,90-93,98%), mientras que la validez de constructo ($n = 36$) demostró una correlación moderada ($r = 0,511-0,517$).

Conclusiones:

La revisión sistemática reveló la ausencia de un cuestionario IPAQ para niños de 5-7 años y la necesidad de investigación de calidad, especialmente para la validez de constructo.

Los estudios sobre propiedades de medición del PAQ-YC demostraron que es un cuestionario con una adecuada validez de contenido, una fiabilidad intraobservador excelente y una validez de contenido moderada.

Palabras clave

Actividad física, cuestionarios, propiedades de medición, pediatría

Abstract

Introduction:

Childhood is a critical period in the consolidation of healthy habits, such as the practice of physical activity. Among the instruments available to measure this construct, questionnaires offer certain advantages, but the lack of high-quality studies makes it difficult to choose the most appropriate one. The International Physical Activity Questionnaire (IPAQ) group is highly recognized within the scientific community and has versions adapted to different populations, countries and cultures.

Methods:

On the one hand, a systematic review on the measurement properties of the IPAQ questionnaires for preschoolers and school children is presented. MEDLINE, Embase and Sport Discus databases were consulted. Two reviewers assessed the methodological quality of the studies using the COSMIN Risk of Bias Checklist. The results were pooled and scored according to the criteria for good measurement properties. The quality of the evidence was scored using the modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE).

On the other hand, the Physical Activity Questionnaire - Young Children (PAQ-YC) for children aged 5-7 years was developed. The first version was tested by a 2-round Delphi study. It was established as a consensus criterion that the Relative Interquartile Range (RIR) and/or the Coefficient of Variation (CV) were $\leq 20\%$. Interviews and a field test were conducted on representative samples of the target population. Next, the reliability and the measurement error were analysed. The questionnaire was administered twice with an interval of 1-3 days. The Kappa coefficient or the Interclass Correlation Index (ICC) as well as the Standard Error of Measurement (SEM) and/or the Percentage of Agreement (PoA) were presented. Finally, the construct

validity was studied by comparing the PAQ-YC with ActiGraph wGT3x-BT accelerometer (ActiGraph Corp., Pensacola, FL). The raw data were processed with Actilife software as vector magnitude measures. The results were presented as Pearson's correlation coefficients.

Results:

21 articles were included in the qualitative analysis and 17 articles were included in the quantitative analysis. The Early Years - Physical Activity Questionnaire (EY-PAQ) for children under 5 years and the Physical Activity Questionnaire - Older Children (PAQ-C) for children aged 8-14 years were identified. Reliability was rated as insufficient for the EY-PAQ (ICC = 0,35-0,47) and as sufficient for the PAQ-C (ICC = 0,79-0,82), both with very low quality of evidence. Construct validity was insufficient in all cases. For the EY-PAQ ($r = 0,02-0,30$) the quality of the evidence was very low, while for the PAQ-C ($r = 0,30-0,52$) it was moderate. Content validity, measurement error, and responsiveness could not be assessed.

For the PAQ-YC, the most significant discrepancies in the Delphi survey ($n = 11-13$) were observed for items 7 (RIR = 20, CV = 38,73) and 8 (RIR = 25, CV = 15,45). The cognitive interviews ($n = 5$) and the field test ($n = 104$) confirmed the version agreed by the experts. Reliability ($n = 83$) showed excellent indices (ICC = 0,983-0,991; SEM = 0,28-2,88; PoA = 75,90-93,98%), while construct validity ($n = 36$) showed a moderate correlation ($r = 0,511-0,517$).

Conclusions:

The systematic review revealed the absence of an IPAQ questionnaire for children aged 5-7 years and the need for quality research, especially for construct validity.

Studies on the measurement properties of the PAQ-YC demonstrated that it is a questionnaire with adequate content validity, excellent intraobserver reliability, and moderate content validity.

Keywords

Physical activity, questionnaires, measurement properties, paediatrics

Prólogo

A lo largo de los años, las ciencias de la actividad física y la epidemiología han liderado la investigación de la actividad física y su impacto en la salud pública. Si bien continúan siendo campos fundamentales en el avance científico de la actividad física, se han sumado nuevas disciplinas que recogen la multidimensionalidad de este concepto, como las ciencias clínicas, las ciencias del comportamiento, la planificación urbana o las ciencias políticas (1,2).

Esta tesis doctoral se enmarca en el estudio de los cuestionarios como instrumentos de medición de los niveles de actividad física en edad preescolar y escolar. Los motivos que han suscitado la elaboración de este trabajo se resumen en tres puntos.

En primer lugar, el estudio de la actividad física como predictor del estado de salud ha adquirido una especial importancia en las últimas décadas (1,3). En 1974, el gobierno canadiense publicó un informe titulado *“A new Perspective on the Health of Canadians”* más comúnmente conocido como *“El Informe Lalonde”* (4). Dentro de los capítulos que trataban la prevención de enfermedades, llamó la atención la importancia que se daba al estilo de vida, incluida la práctica de actividad física. El informe canadiense fue seguido por el informe del gobierno de EE.UU., en el que se repetía el mensaje acerca de los beneficios de la práctica habitual de actividad física (5). Ambos documentos fueron pioneros en destacar el impacto de las conductas de estilo de vida sobre el riesgo de la enfermedad, una afirmación que a día de hoy está claramente descrita y aceptada por la comunidad científica (1,3).

En segundo lugar, se reconoce que la infancia es un período clave en el desarrollo y consolidación de hábitos saludables. Durante esta etapa, el individuo es sensible a ser influenciado por el medio ambiente en sus conductas modificables, como la práctica habitual de actividad física (6).

En tercer lugar, se ha documentado un cuerpo de evidencia pobre en relación a los cuestionarios de medición de la actividad física en etapa preescolar y escolar (7-13). Dado que este método sería un medio viable para medir los niveles de actividad de la población, es necesario desarrollar y validar adecuadamente nuevos instrumentos de informes indirectos que apoyen el estudio de la actividad física desde edades tempranas (14,15).

Durante los próximos capítulos se abordarán aspectos relacionados con la terminología, el marco conceptual y otras cuestiones prácticas relacionadas con la medición de la actividad física. En primer lugar, se explicarán las diferencias en la terminología y definiciones de los conceptos comúnmente empleados en el estudio de la actividad física (Capítulo 1.1). A continuación, se profundizará en el marco conceptual que subyace al comportamiento del movimiento humano como base para estandarizar la medición de la actividad física (Capítulo 1.2). Por último, se abordarán los instrumentos de medición disponibles, especialmente enfocado hacia el estudio de los cuestionarios de actividad física dirigidos a niños en edad preescolar y escolar (Capítulo 1.3).

Índice de contenidos

1.	Introducción.....	1
1.1.	Taxonomía y definiciones	1
1.2.	Marco conceptual de la actividad física	5
1.2.1.	Influencia de los factores internos y externos.....	6
1.2.2.	Dominios de la actividad física y comportamiento sedentario	10
1.2.2.1.	Tiempo libre.....	11
1.2.2.2.	Escuela.....	17
1.2.2.3.	Transporte	20
1.2.3.	Desarrollo de los patrones del movimiento humano	20
1.2.4.	Atributos fisiológicos	24
1.2.4.1.	Gasto energético.....	24
1.2.4.2.	Efectos de la actividad física y el comportamiento sedentario	30
1.3.	Métodos de medición de la actividad física y el gasto energético	32
1.3.1.	Instrumentos de medición del gasto energético	34
1.3.2.	Instrumentos de medición de la actividad física	37
1.3.3.	¿Cómo escoger un instrumento de medición?.....	42
1.3.4.	¿Cómo escoger un cuestionario de actividad física?.....	47
1.3.4.1.	Atributos cualitativos	48
1.3.4.1.1.	Constructo.....	48
1.3.4.1.2.	Población diana	50
1.3.4.1.3.	Período de recuerdo.....	51
1.3.4.1.4.	Formato.....	52
1.3.4.1.5.	Objetivo	53
1.3.4.1.6.	Interpretabilidad.....	54

1.3.4.1.7. Limitaciones logísticas	55
1.3.4.2. Propiedades de medición.....	56
1.3.4.2.1. Validez de contenido	62
1.3.4.2.2. Validez de criterio	64
1.3.4.2.3. Validez de constructo.....	64
1.3.4.2.4. Consistencia interna.....	68
1.3.4.2.5. Fiabilidad.....	69
1.3.4.2.6. Error de medición	70
1.3.4.2.7. Capacidad de respuesta	71
1.3.4.3. Cuestionarios dirigidos a edad preescolar y escolar	72
2. Justificación	76
3. Cuestionarios internacionales de actividad física en edad escolar y preescolar: una revisión sistemática de propiedades de medición	80
3.1. Objetivos generales.....	81
3.2. Objetivos específicos.....	82
3.3. Material y métodos	83
3.3.1. Criterios de elegibilidad.....	83
3.3.2. Fuentes de información y búsqueda.....	84
3.3.3. Selección de los artículos	90
3.3.4. Evaluación de la calidad metodológica de los estudios.....	90
3.3.5. Aplicación de los criterios para buenas propiedades de medición	92
3.3.6. Resumen de la evidencia por propiedad de medición.....	93
3.3.7. Calificación de la calidad de la evidencia	94
3.4. Resultados.....	96

3.4.1.	Validez de contenido	98
3.4.2.	Validez transcultural	100
3.4.3.	Fiabilidad.....	103
3.4.4.	Error de medición	106
3.4.5.	Validez de constructo.....	107
3.5.	Discusión	117
3.5.1.	Validez de contenido	118
3.5.2.	Validez transcultural.....	118
3.5.3.	Fiabilidad.....	125
3.5.4.	Error de medición	129
3.5.5.	Validez de constructo.....	130
3.5.6.	Puntos fuertes y limitaciones.....	137
3.5.7.	Consideraciones generales.....	137
4.	Desarrollo y propiedades de medición del Physical Activity Questionnaire - Young Children (PAQ-YC)	139
4.1.	Hipótesis general	140
4.2.	Hipótesis específicas	141
4.3.	Objetivos generales.....	142
4.4.	Objetivos específicos	143
4.5.	Material y métodos	144
4.5.1.	Desarrollo y validez de contenido.....	144
4.5.1.1.	Descripción del constructo.....	144
4.5.1.2.	Selección de los ítems	145
4.5.1.3.	Elaboración del sistema de puntuación.....	147
4.5.1.4.	Consultas a expertos y a muestras de la población diana	155

4.5.1.5.	Prueba de campo	158
4.5.2.	Fiabilidad y error de medición	159
4.5.3.	Validez de constructo.....	160
4.5.4.	Consideraciones éticas	162
4.6.	Resultados.....	164
4.6.1.	Desarrollo y validez de contenido.....	164
4.6.1.1.	Prueba de campo	179
4.6.2.	Fiabilidad y error de medición	180
4.6.3.	Validez de constructo.....	183
4.7.	Discusión	188
4.7.1.	Desarrollo y validez de contenido.....	188
4.7.2.	Fiabilidad y error de medición	210
4.7.3.	Validez de constructo.....	213
4.7.4.	Puntos fuertes y limitaciones.....	219
4.7.5.	Consideraciones generales.....	221
5.	Conclusiones	223
6.	Líneas de futuro.....	226
7.	Referencias bibliográficas.....	228
8.	Anexos.....	277
8.1.	Anexo 1. Riesgo de sesgo de COSMIN. Versión 2018.....	278
8.2.	Anexo 2. Versión original del EY-PAQ.....	284
8.3.	Anexo 3. Versión original del PAQ-C.....	286
8.4.	Anexo 4. Aprobación del comité de ética de investigación.....	289
8.5.	Anexo 5. Documento de información y consentimiento informado del estudio de fiabilidad	290

8.6.	Anexo 6. Documento de información y consentimiento informado del estudio de validez de constructo.....	293
8.7.	Anexo 7. Development and Content Validity of the Physical Activity Questionnaire-Young Children (PAQ-YC) to Assess Physical Activity in Children between 5 and 7 Years.....	297

Índice de tablas

Tabla 1. Tiempo semanal dedicada a las clases de Educación Física según el nivel escolar. España 2019. Estudio ALADINO	19
Tabla 2. Definiciones COSMIN de las propiedades de medición. Adaptación de Mokkink et al. 2010.....	61
Tabla 3. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios de actividad física. Adaptación de Terwee et al. 2010.	65
Tabla 4. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios de actividad física. Niveles de evidencia 1, 2 y 3. Adaptación de Hidding 2018	66
Tabla 5. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios que miden el comportamiento sedentario. Niveles de evidencia 1, 2 y 3. Adaptación de Hidding 2017	67
Tabla 6. Estrategia de búsqueda en PubMed, EMBASE y Sport Discus.....	86
Tabla 7. Características generales de los artículos de validación original y de validez transcultural.....	101
Tabla 8. Estudios de fiabilidad. Evaluación de la calidad metodológica y calificación del resultado	103
Tabla 9. Meta-análisis de todos los estudios de fiabilidad del PAQ-C	104
Tabla 10. Meta-análisis de los estudios de fiabilidad del PAQ-C que han obtenido una calificación de la calidad metodológica de al menos "dudosa".	105
Tabla 11. Estudios de error de medición del PAQ-C Evaluación de la calidad metodológica y calificación del resultado	107

Tabla 12. Estudios de validez de constructo. Evaluación de la calidad metodológica y calificación del resultado	111
Tabla 13. Meta-análisis de los estudios de validez de constructo que no aportan evidencia de suficiente calidad.....	113
Tabla 14. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 3	113
Tabla 15. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2	114
Tabla 16. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos uniaxiales.....	115
Tabla 17. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos triaxiales.....	115
Tabla 18. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 1	116
Tabla 19. Equivalencias METs para los ítems 1, 2, 3 y 4 del PAQ-YC.....	150
Tabla 20. Equivalencias de puntuación para los ítems 1, 2, 3 y 4 según los MET-min/día.....	152
Tabla 21. Puntuación del ítem 8 con factor de corrección del ítem 7.....	154
Tabla 22. Puntuación de los ítems 9 y 10.....	154
Tabla 23. Ponderación del Coeficiente de Conocimiento (Kc)	156
Tabla 24. Grado de acuerdo para cada uno de los ítems en las rondas 1 y 2	170

Tabla 25. Grado de acuerdo para cada una de las características generales en las rondas 1 y 2.....	171
Tabla 26. Distribución de las puntuaciones del PAQ-YC en la prueba de campo	180
Tabla 27. Distribución de las puntuaciones del PAQ-YC en el estudio de fiabilidad y error de medición.....	181
Tabla 28. Fiabilidad y error de medición de los ítems que contribuyen a la puntuación del PAQ-YC	182
Tabla 29. Fiabilidad y error de medición de la puntuación del Bloque 1, Bloque 2, Total, porcentaje de actividad física y porcentaje de comportamiento sedentario discrecional	183
Tabla 30. Distribución de las puntuaciones del PAQ-YC en el estudio de validez de constructo.....	184
Tabla 31. Distribución de los registros de los dispositivos Actigraph wGT3x-BT (ActiGraph Corp., Pensacola, FL)	185
Tabla 32. Validez de constructo por comparación de métodos entre el PAQ-YC y acelerometría Actigraph wGT3X-BT.....	186

Índice de figuras

Figura 1. Relación entre actividad física moderada-vigorosa, comportamiento sedentario y riesgo de mortalidad. Adaptación de Ekelund et al. 2016	4
Figura 2. Esquema del marco conceptual del comportamiento del movimiento humano. Adaptación de Pettee Gabriel KK et al. 2012.....	5
Figura 3. Perspectiva socio ecológica de los niveles de influencia de la actividad física	7
Figura 4. Tiempo libre frente a una pantalla según clase socioeconómica basado en la ocupación de la persona de referencia (%). España 2017. ENSE, MSCBS/INE	9
Figura 5. Tiempo libre frente a una pantalla según grupo de edad (%). España 2017. ENSE, MSCBS/INE.....	13
Figura 6. Tiempo dedicado a realizar los deberes o leer libros. España 2019. Estudio ALADINO	14
Figura 7. Tiempo libre frente a una pantalla los días entre semana y el fin de semana. España 2019. Estudio ALADINO	14
Figura 8. Tiempo libre dedicado a actividades deportivas extraescolares. España 2019. Estudio ALADINO.....	16
Figura 9. Tiempo libre dedicado a practicar actividad física de intensidad moderada a vigorosa. España 2019. Estudio ALADINO.....	16
Figura 10. Modelo conceptual de los métodos de medición de la actividad física y el gasto energético. Adaptación de Lamonte et al. 2001	33
Figura 11. Algoritmo de decisión. Adaptación de The European Association of Cardiovascular Prevention and Rehabilitation. 2010	45

Figura 12. Algoritmo de decisión. Adaptación de Pettee Gabriel KK et al. 2012	45
Figura 13. Algoritmo de decisión. Adaptación de The American Heart Association. 2013.....	46
Figura 14. Proceso de recogida de información en los cuestionarios de actividad física. Adaptación de Pettee Gabriel KK et al. 2012.....	49
Figura 15. Representación de un modelo reflexivo y un modelo formativo. Adaptación de Vet et al. 2011	57
Figura 16. Taxonomía COSMIN de las propiedades de medición. Adaptación de Mokkink et al. 2010.....	60
Figura 17. Diagrama de flujo PRISMA para la selección de los artículos incluidos.....	97
Figura 18. Forest-plot del meta-análisis de todos los estudios de fiabilidad del PAQ-C.....	105
Figura 19. Forest-plot del meta-análisis de los estudios de fiabilidad del PAQ-C que han obtenido una calificación de la calidad metodológica de al menos "adecuada"	106
Figura 20. Forest-plot de los estudios de validez de constructo que no aportan evidencia de suficiente calidad	113
Figura 21. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 3.....	114
Figura 22. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2.....	114

Figura 23. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos uniaxiales	115
Figura 24. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos triaxiales	116
Figura 25. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 1	116
Figura 26. Formato de las opciones de respuesta del PAQ-YC	147
Figura 27. Ítems 1, 2, 3, y 4 del PAQ-YC	149
Figura 28. Ítems 5 y 6 del PAQ-YC	151
Figura 29. Porcentajes de respuesta para cada uno de los ítems y las características generales	169
Figura 30. Versión final del PAQ-YC	178
Figura 31. Gráficos de dispersión entre las variables analizadas para la validez de constructo del PAQ-YC.....	187

Índice de abreviaturas

ACSM	American College of Sports Medicine
AF	Actividad Física
AFC	Análisis Factorial Confirmatorio
AFMV	Actividad Física Moderada-Vigorosa
AHA	American Heart Association
APA	Asociación Americana de Psicología
ATN	Active Transportation to School and Work in Norway
CAFÉ	Ciencias de la Actividad Física y el Deporte
CAR	Computerized Activity Recall
CHMS	Canadian Health Measures Survey
CLASS	Children's Leisure Activities Study Survey
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CPAQ	Children's Physical Activity Questionnaire
cpm	recuentos por minuto
CS	Comportamiento Sedentario
CV	Coefficiente de Variación
DIF	Funcionamiento Diferencial de Ítems
EFA	Análisis Factorial Exploratorio
ERBs	Energy Balance Related Behaviours
FC	Frecuencia Cardíaca
GAQ	Girls Health Enrichment Multisite Study Activity Questionnaire
GE	Gasto Energético
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
ICC	Índice de Concordancia Intraobservador
INE	Instituto Nacional de Estadística

IPAQ	International Physical Activity Questionnaire
IRT	Teoría de Respuesta al Ítem
I ²	Coeficiente de Inconsistencia
Kc	Coeficiente de Conocimiento
LoA	Límites del Acuerdo
LOMCE	Ley Orgánica para la Mejora de la Calidad Educativa
LOMLOE	Ley Orgánica de Modificación de la Ley Orgánica de Educación
MARCA	Multimedia Activity Recall for Children and Adolescents
MET	Equivalente Metabólico
NPAQ	Netherlands Physical Activity Questionnaire for Young Children
OMS	Organización Mundial de la Salud
PAI	Physical Activity Interview
PAQ-C	Physical Activity Questionnaire – Older Children
PAQ-YC	Physical Activity Questionnaire - Young Children
PDPAR	Previous Day Physical Activity Recall
PoA	Porcentaje de Acuerdo
Pre-PAQ	Preschool-age Children’s Physical Activity Questionnaire
PRISMA	Preferred Items for Systematic Reviews and Meta-Analyses
Q1	Cuartil 1
Q3	Cuartil 3
RIR	Rango Intercuartil Relativo
SAC-MOS	Comité Asesor Científico del Medical Outcomes Trust
SAYCARE	South American Youth/Child Cardiovascular and Environment Study
SD	Desviación Estándar
SDC	Cambio más pequeño detectable
SEM	Error Estándar de Medición
SRHI	Self-Report Habit Index
TMR	Tasa Metabólica en Reposo

VCO ₂	Producción de dióxido de carbono
VO ₂	Consumo de oxígeno
YACH	Yesterday Activity Checklist
7-DPAR	7 - Days Physical Activity Recall

1. Introducción

1.1. Taxonomía y definiciones

La definición de actividad física (AF) más reconocida a nivel internacional fue publicada en 1985 por Caspersen et al. (2) en un intento de estandarizar la terminología utilizada para referirse a los conceptos de “actividad física”, “ejercicio” y “aptitud o condición física”. Aunque no es objeto de esta tesis profundizar en la taxonomía y definiciones, creemos conveniente detenernos un momento para aclarar los matices que recoge cada concepto.

Caspersen et al. definieron la AF como *“cualquier movimiento corporal producido por los músculos esqueléticos que resulta en un gasto de energía”* (2). Con el paso de los años, se han propuesto otras definiciones, sin embargo, la mayoría son derivaciones de la definición de Caspersen. Por ejemplo, en 1995, la declaración de consenso de los Institutos Nacionales de Salud de EE.UU. añadió en la página 3 de su informe que la AF *“...produce beneficios para la salud”* (16). Un año más tarde, el Departamento de Salud y Servicios Humanos en la página 20 de su manuscrito aclaró que la AF incluye el gasto de energía *“...por encima del nivel basal”*, mientras que en la siguiente página se refiere a la AF como cualquier actividad que *“aumenta sustancialmente el gasto energético”* (17). En 2018, la estrategia mundial sobre AF de la Organización Mundial de la Salud (OMS) matizó que la AF es el movimiento corporal que *“...requiere gasto energético”* (18). Todas estas definiciones interpretan la AF desde la perspectiva de la epidemiología relacionada con la salud y ponen el foco en el movimiento corporal, los músculos esqueléticos, el gasto energético (GE) o la aptitud física.

La razón por la que la definición Caspersen et al. (2) sigue siendo la más aceptada por la comunidad científica es su simplicidad y claridad. Por ejemplo, la definición de los Institutos Nacionales de Salud de EE.UU. (16) ha sido criticada porque las actividades como el esfuerzo repetitivo se contabilizan como AF, pero no necesariamente producen beneficios para la

salud (19). La propuesta del Departamento de Salud y Servicios Humanos de EE.UU. (17) tampoco ha tenido éxito ya que asocia la AF a las formas de intensidad moderada a vigorosa cuando en realidad debe recoger la gama completa de intensidades, incluidas las actividades de intensidad ligera (3).

Fruto de la gran variedad de disciplinas involucradas en el estudio de la AF, recientemente se ha propuesto una nueva definición más inclusiva y holística que contempla otros aspectos inherentes a la actividad. Esta definición fue publicada en 2020 y recoge que *“la actividad física involucra el movimiento de las personas, actuando y desenvolviéndose dentro de espacios y contextos culturalmente específicos e influenciados por una variedad de intereses, emociones, ideas, instrucciones y relaciones”*. Como novedad, redirige la atención a la persona en lugar de a los músculos esqueléticos e incluye los componentes psicológicos, sociales, contextuales y políticos dentro de la propia definición (19). Al ser de tan reciente publicación, todavía no ha sido objeto de debate entre la comunidad científica.

Aunque en ocasiones se utilizan indistintamente los términos “actividad física” y “ejercicio”, en realidad el ejercicio es una subcategoría de la AF, definida como *“una actividad física planificada, estructurada, repetitiva y con el objetivo de mejorar o mantener uno o más componentes de la aptitud o condición física”* (2). El matiz entre ambos conceptos radica en la planificación y el objetivo de la actividad, pudiendo una misma actividad ser considerada AF o ejercicio en función del contexto. Por ejemplo, las actividades domésticas habitualmente se realizan de la forma más eficiente posible priorizando la conservación del GE e ignorando la aptitud física. En este caso, deben considerarse AF. Por el contrario, uno puede planificar y estructurar el desempeño de dichas actividades de una manera menos eficiente para desarrollar fuerza muscular o incrementar el GE. En este caso, las tareas domésticas se considerarían ejercicio (20).

A diferencia de la AF, la “aptitud o condición física” es un conjunto de atributos definidos como *“la capacidad de realizar las tareas diarias con vigor y alerta, sin fatiga indebida y con la energía suficiente para disfrutar de las actividades de ocio y hacer frente a situaciones imprevistas que requieren un esfuerzo complementario”*. Estos atributos incluyen la resistencia cardiorrespiratoria, la resistencia muscular, la fuerza muscular, la flexibilidad, el equilibrio, la coordinación y la composición corporal, todos ellos ampliamente descritos como indicadores del estado de salud (2,20). Aunque la definición puede ser conceptualmente sólida, aspectos como el vigor, el estado de alerta, la fatiga y el disfrute son conceptos abstractos difíciles de medir. Sin embargo, la idea que representa la aptitud física no es una subcategoría de la AF y genera menos confusión que en el caso anterior. La aptitud física se refiere a la capacidad del individuo para realizar la AF y, aunque son conceptos que están íntimamente relacionados, la aptitud física no engloba a la AF, sino que permite que esta se realice con eficiencia (2).

Al examinar el perfil conductual del movimiento humano, también se debe considerar el comportamiento sedentario (CS). Durante años, el CS se ha identificado como lo opuesto a la AF, entendiendo que aquel individuo que era inactivo físicamente, en consecuencia, era un individuo sedentario. Sin embargo, en la última década, se ha demostrado que son comportamientos independientes y se ha reconocido que la actividad (o inactividad) física y el CS pueden combinarse e incluso coexistir. Por ejemplo, es posible que un individuo cumpla las recomendaciones de práctica de AF semanal y simultáneamente presente un perfil sedentario por pasar más de 8 horas diarias sentado (por ejemplo, en las horas de escuela y el estudio en casa). Del mismo modo, si una persona se considera poco activa al cuantificar sus niveles de AF, no significa necesariamente que sea también sedentaria (21).

En 2012, la Red de Investigación del Comportamiento Sedentario publicó una carta para aclarar las diferencias entre CS e “inactividad física”. En ella

proponía que se adoptara una única descripción del CS, definida como *“cualquier conducta de vigilia caracterizada por un gasto de energía $\leq 1,5$ METs en una postura sentada o reclinada”*. Por el contrario, sugería utilizar el término *“inactividad física”* para identificar a aquellos individuos que *“realizan cantidades insuficientes de actividad física moderada-vigorosa, es decir, no cumplen con las pautas específicas de actividad física”* (22).

En 2016, Ekelund et al. publicaron un estudio que ilustra esta relación. Sus hallazgos demostraron que la AF y el CS son comportamientos independientes y documentaron su aportación individual a la predicción del riesgo de mortalidad por enfermedad cardiovascular (Figura 1). En el mapa de color se observa que incluso para aquellos individuos altamente sedentarios, la adición de una dosis adecuada de AF disminuye el riesgo de mortalidad, pudiendo ser muy bajo si los niveles de AF son óptimos. Sin embargo, los niveles bajos de CS pueden limitar el riesgo de mortalidad por causa cardiovascular, pero no consiguen reducirlo al mínimo a no ser que se acompañen de una práctica regular de AF de intensidad moderada a vigorosa (AFMV) (23).



Figura 1. Relación entre actividad física moderada-vigorosa, comportamiento sedentario y riesgo de mortalidad. Adaptación de Ekelund et al. 2016 (23)

1.2. Marco conceptual de la actividad física

Para avanzar en el estudio de la AF, es necesario describir y utilizar un marco de referencia que proporcione la base para estandarizar el desarrollo de una investigación homogénea y de calidad. Aunque se han propuesto varios marcos conceptuales, la amplia variedad de disciplinas que estudian el comportamiento de la AF ha hecho que coexistan varios enfoques que tratan la AF desde diferentes perspectivas (24–27).

En 2012, Pettee Gabriel KK et al. definieron un marco teórico basado en el comportamiento humano que implica una relación direccional entre el aspecto conductual, la característica del movimiento y el resultado o consecuencia fisiológica (Figura 2). Estos autores clasificaron el comportamiento en conductas activas y sedentarias, reconociendo que ambas pueden verse influidas por factores fisiológicos, psicológicos, sociales y ambientales. Por último, señalaron que la consecuencia directa del comportamiento activo es la producción de movimiento, que da como resultado atributos fisiológicos que incluyen un mayor gasto de energía y una mejor condición física (20).

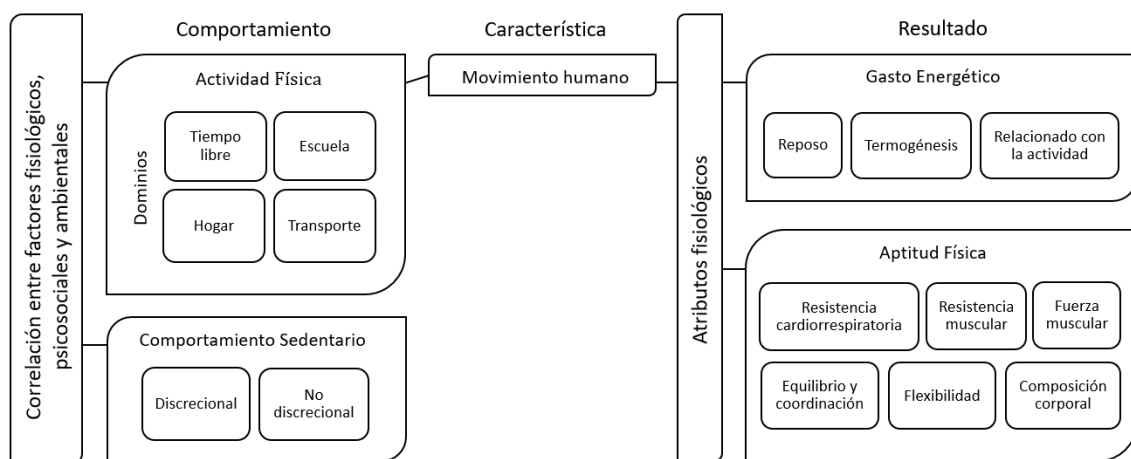


Figura 2. Esquema del marco conceptual del comportamiento del movimiento humano. Adaptación de Pettee Gabriel KK et al. 2012 (20)

1.2.1. Influencia de los factores internos y externos

La perspectiva del marco conceptual de la AF propuesto por Pettee Gabriel KK et al. se fundamenta en la teoría de sistemas dinámicos descrita por Thelen y Smith en 1994. La teoría de sistemas dinámicos propone que el desarrollo del comportamiento infantil está influenciado por sistemas y subsistemas interactivos, tanto internos como externos al individuo. Por tanto, la AF y el CS, como parte del comportamiento humano, están claramente asociados a factores fisiológicos, psicosociales y ambientales (Figura 1) (28).

A lo largo de los años se ha descrito que la herencia genética contribuye al fenotipo de actividad (o inactividad) y se han identificado diferentes mecanismos genéticos, neuroendocrinos y materno-fetales involucrados en la predisposición a la AF (29,30). Las teorías clásicas sobre el comportamiento de la AF daban mucha importancia al estudio de los factores internos, con un reconocimiento modesto a las influencias externas. Sin embargo, se ha demostrado que los factores genéticos apenas tienen peso en la tendencia del individuo a ser activo y, por el contrario, el entorno y los factores interpersonales son los condicionantes más importantes, representando alrededor del 70% (31) y del 20-40% (32) de la varianza atribuible a este fenotipo, respectivamente. Por tanto, la investigación se ha desplazado hacia el uso de modelos socio ecológicos que postulan que el comportamiento es multidimensional y está influenciado principalmente por la interacción de factores personales, socioculturales y el entorno físico (Figura 3) (33,34).

Dentro de los factores personales, la teoría de la autodeterminación enfatiza la importancia de la percepción de autonomía, competencia percibida y conexión social. La teoría postula que las percepciones personales son una necesidad psicológica fundamental que influyen en el bienestar y promueven formas autónomas de motivación, que a su vez predicen el compromiso y la persistencia conductual (35). La percepción de los niños de que son capaces

de realizar con éxito una tarea se relaciona con la participación repetida en dicha actividad. Por ejemplo, durante los primeros años de escuela, la autodeterminación depende de la realización satisfactoria de tareas sencillas (por ejemplo, atrapar una pelota) y la retroalimentación para construir una percepción de competencia (36).

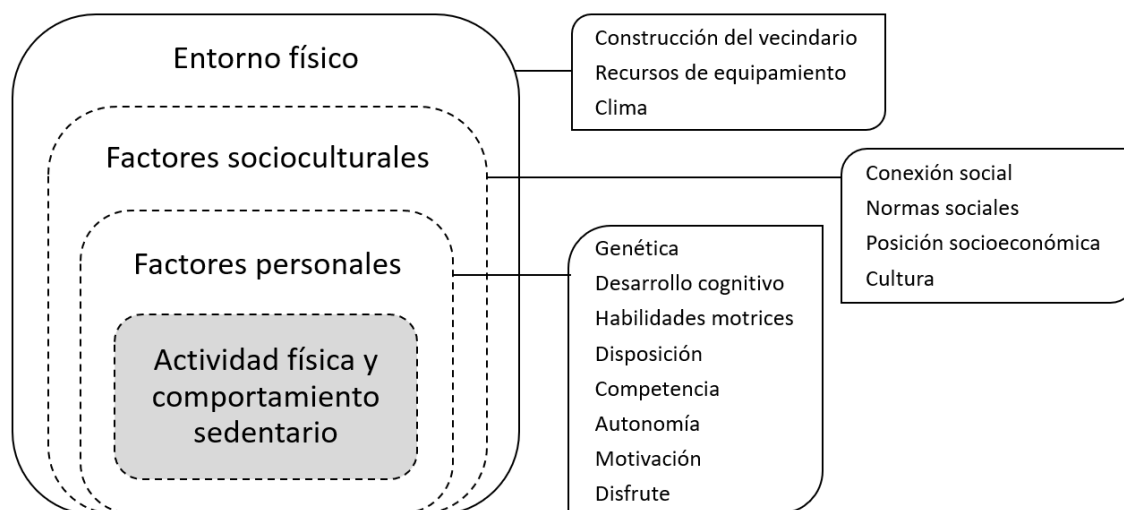


Figura 3. Perspectiva socio ecológica de los niveles de influencia de la actividad física

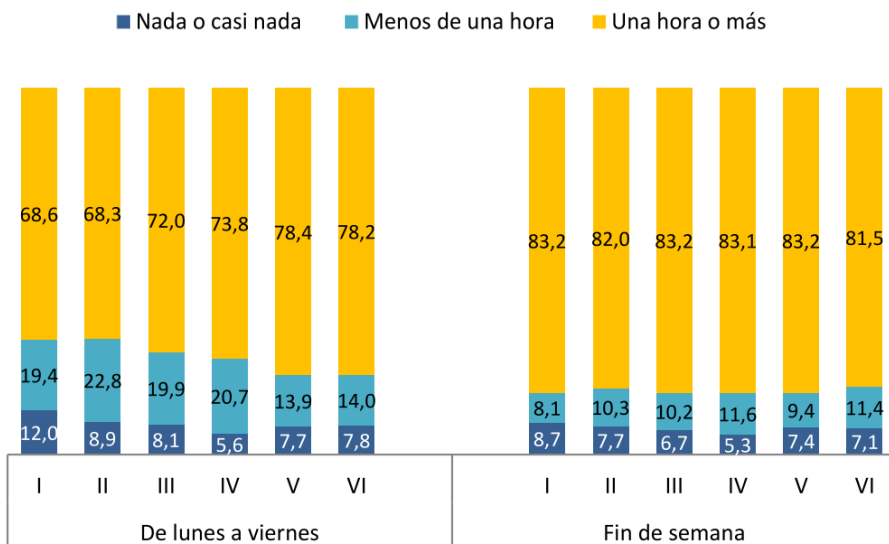
Otro de los predictores más fuertes de la participación en AF es el disfrute. El principio de disfrute se alinea con las teorías de la motivación que defienden que las personas tienden a persistir en las actividades que encuentran atractivas en sí mismas (37). Cuando las personas realizan AF (o cualquier otra actividad) por la alegría y el placer inherentes, se dice que están intrínsecamente motivadas y presentan una mayor adherencia y repetición de ese comportamiento (38). En el contexto de la AF en la infancia, cuando a los niños se les brinda la oportunidad de elegir la modalidad preferida, tienden a mostrar un mayor compromiso y una mayor intención de continuar con dicha actividad en el futuro (39).

Desde la perspectiva de las teorías socio ecológicas, se recoge que la conducta también está determinada por factores sociales, como la influencia de los

maestros, los padres y los compañeros. En general, las conductas del entorno pueden influir en la motivación apoyando o frustrando las necesidades psicológicas básicas de autonomía, competencia y apoyo social (37). Por ejemplo, las normas sociales pueden determinar la actitud de un niño hacia una AF por su percepción de aceptabilidad social como una actividad inusual o común. En definitiva, la práctica de una determinada actividad no solo depende de la visión personal, sino que es un fenómeno social compartido, aprendido e influenciado por otros (24).

Además, existe una relación entre la posición socioeconómica y el nivel de AF durante la infancia y como predictor en la edad adulta (40). En general, los niños que provienen de familias con un nivel socioeconómico más bajo utilizan con mayor frecuencia medios basados en pantalla en su tiempo libre y son menos activos físicamente (41). En la escuela, se ha descrito que los efectos de agrupamiento, por ejemplo, de un mayor porcentaje de alumnos de baja posición económica o de minorías étnicas puede constituir un factor contextual que condicione la AF (42,43). Las escuelas ubicadas en barrios de alto nivel socioeconómico disponen de los medios para brindar ambientes más propicios para la AF y se asocian a una mayor presencia de conductas activas durante el horario lectivo (44).

En el 2017, el Ministerio de Sanidad, Consumo y Bienestar Social realizó una encuesta nacional para conocer los hábitos nutricionales y de AF en España. Con respecto al uso recreativo de medios de pantalla, describieron que los días lectivos, el 78,2% de los menores de clases poco favorecidas pasaba más de 1 hora con dispositivos electrónicos, frente a un 68,6% de los niños de clase socioeconómica alta (Figura 4) (45). Estos hallazgos sugieren la necesidad de proporcionar oportunidades y apoyo adicionales para desarrollar un ocio y una escolarización más activos a los niños en situación de desventaja socioeconómica que además podrían ser un medio para debilitar el vínculo entre los sectores infantiles más desfavorecidos (40).



I-VI: status socioeconómico de menor a mayor nivel

Figura 4. Tiempo libre frente a una pantalla según clase socioeconómica basado en la ocupación de la persona de referencia (%). España 2017. ENSE, MSCBS/INE (45)

Por otro lado, se ha descrito una clara asociación entre los niveles de AF practicados por la población y la construcción del entorno físico del vecindario (46). Específicamente, se ha demostrado que los ambientes caracterizados como más transitables (mayor proporción de trayectos peatonales, conectividad en las aceras, calidad de la infraestructura, etc.) se relacionan con la acumulación de minutos de AF durante los desplazamientos (47). Una revisión sistemática del 2017 demostró también un efecto positivo de la provisión de parques y áreas de juego de calidad en la práctica de AF durante el tiempo libre. Además, informaron que la construcción de entornos facilitadores de conductas activas mejora el paisaje urbano y fomenta la creación de un entorno en el que es fácil combinar tanto espacios lúdicos como de transporte activo (48). El estudio ALADINO del año 2019 demostró que, en España, el 65,3% de las familias disponen de un polideportivo o zona verde a menos de 1 km de su hogar, mientras que un 23,4% lo tiene entre 1 y 2 km de distancia (49).

Por último, hay que considerar que la AF depende en gran medida del clima, especialmente en países donde las estaciones del año son más extremas, ya que puede variar la posibilidad de participar en actividades al aire libre y modificar drásticamente las actividades habituales en período de vacaciones. El conocimiento de las variaciones estacionales en el comportamiento de la salud de los niños facilita una orientación más precisa de las intervenciones de cambio de comportamiento y también es importante en el contexto de la evaluación ya que las estimaciones pueden estar sesgadas según el período del año en el que sean recogidas (50).

1.2.2. Dominios de la actividad física y comportamiento sedentario

Pettee Gabriel KK et al. identificaron cuatro dominios (tiempo libre, trabajo, tareas domésticas o de cuidado personal y transporte) que representan los escenarios para realizar AF en la vida diaria. Además, destacan la importancia de considerar el CS como una característica de la conducta humana de gran interés en el campo de la salud pública (20,21). Proponen clasificar el CS en discrecional y no discrecional, haciendo hincapié en la influencia social y/o comunitaria de la conducta. De esta forma, distinguen las actividades sedentarias no modificables por el individuo (como el tiempo sentado en la escuela o en el trabajo) de aquellas que el individuo escoge libremente (como el tiempo de ocio con dispositivos de pantalla) (Figura 2) (20). El esquema propuesto por estos autores puede llevar a confusión si entendemos que los dominios son exclusivos de la AF. Es evidente que en todos los contextos se pueden dar tanto conductas activas como sedentarias y un reconocimiento de ambos comportamientos en cada uno de los escenarios supondrá una representación más veraz de la realidad (51-55).

Por otro lado, estas categorías resumen los entornos en los que se puede desarrollar la AF o el CS en el adulto, pero no se adaptan a las actividades

habituales en la infancia. En primer lugar, el dominio tiempo libre en los niños representa una mayor fracción del día (56). En segundo lugar, las labores domésticas o de cuidado personal no tienen gran representación en la vida diaria del niños, entendiendo que a edades tempranas estas tareas corren a cargo de los padres o los cuidadores. En tercer lugar, el dominio trabajo puede sustituirse por el tiempo en la escuela, aunque puede no estar representado en todos los casos. Con respecto a este último punto, en España la escolarización no es obligatoria hasta los 6 años. Según los datos del Instituto Nacional de Estadística (INE) del año 2017, la tasa de escolarización en este período opcional aumenta desde los dos primeros años (40,5% y 60,5% con 1 y 2 años, respectivamente) hasta ser muy próxima al 100% cuando se alcanza el segundo ciclo de Educación Infantil (96,6%, 97,5% y 98,3% con 3, 4 y 5 años, respectivamente) (57).

Comprender la representación de cada uno de los dominios es fundamental en la construcción de instrumentos de medición apropiados. Históricamente, los enfoques para promover la AF y reducir el CS estaban centrados en el dominio tiempo libre y los instrumentos para su evaluación fueron desarrollados y validados en consecuencia. Sin embargo, como el foco de mejora puede incidir en cualquiera de los dominios mencionados, es imprescindible evaluar cada uno de ellos (58).

1.2.2.1. Tiempo libre

El tiempo libre es un descriptor amplio de las actividades en las que un individuo participa voluntariamente basándose en los intereses y necesidades personales. Aunque la tipología de estas actividades abarca un amplio rango de intensidades, el tiempo libre constituye el escenario con mayor potencial para la AFMV (59). Se ha demostrado que el ocio representa del 43% al 50% de la variabilidad en el cómputo total de la AF durante la

infancia (60). Aunque es un porcentaje importante, los niños que no son lo suficientemente activos durante el tiempo libre tienen la oportunidad de acumular minutos adicionales en otros dominios para alcanzar las recomendaciones de AF. A pesar de esta posibilidad, dada la situación actual de las rutinas y los horarios de los padres, esta compensación no suele ocurrir de forma espontánea (51).

El entorno en el que se desarrollan las actividades de ocio incluye el hogar y la comunidad, tanto en instalaciones interiores como en espacios abiertos. Los estudios sobre las conductas en el tiempo libre coinciden en que el ocio en el interior tiene más posibilidades de ser sedentario, mientras que el juego en exteriores es más probable que sea activo. La falta de espacios abiertos, como un patio o un jardín limita las oportunidades de AF en el hogar debido a la fuerte correlación observada entre los niveles de AF y el juego al aire libre. Además, se ha demostrado que el cuidado infantil en las instalaciones interiores del hogar es bastante propicio para las actividades sedentarias (51,52).

El tiempo de ocio está representado principalmente por las horas después de la escuela y el fin de semana. Durante este período, los niños pueden tener más opciones sobre las conductas que realizan en comparación con otros momentos (54). Existen varios estudios que investigan la relación entre diferentes actividades de ocio, especialmente entre el uso recreativo de medios de pantalla y la AF. Se ha reportado alguna evidencia de que los niños físicamente activos recurren menos a actividades de tipo sedentarias y viceversa (41). Sin embargo, también se ha descrito que el uso de dispositivos electrónicos y la práctica de AF pueden coexistir en lugar de competir (61).

Dentro de los CS más frecuentes destacan las actividades basadas en pantalla, como ver la televisión o jugar a videojuegos, aunque también se deben contemplar otros comportamientos diferentes como la lectura, las manualidades o los deberes (41,54).

Las guías sobre el CS recomiendan que los niños menores de 4 años no pasen más de 1 hora diaria frente a una pantalla (62), mientras que para los niños más mayores y los adolescentes se permite un tiempo máximo de 2 horas (63). En el 2017, el Ministerio de Sanidad, Consumo y Bienestar Social realizó una encuesta nacional para conocer los hábitos nutricionales y de AF en España. Con respecto a la población infantil, informaron que el 73,9-82,6% (entre semana y fin de semana, respectivamente) de los niños españoles pasa más de 1 hora al día frente a una pantalla de uso recreativo. Describieron también que la proporción de horas aumenta con la edad, observándose cifras del 57,8-62,9% entre los menores de 4 años que ascienden al 83,7-92,6% en la población de 10 a 14 años. Además, reportaron que el 14% de los niños ocupan su tiempo libre de forma casi totalmente sedentaria (Figura 5)(45).

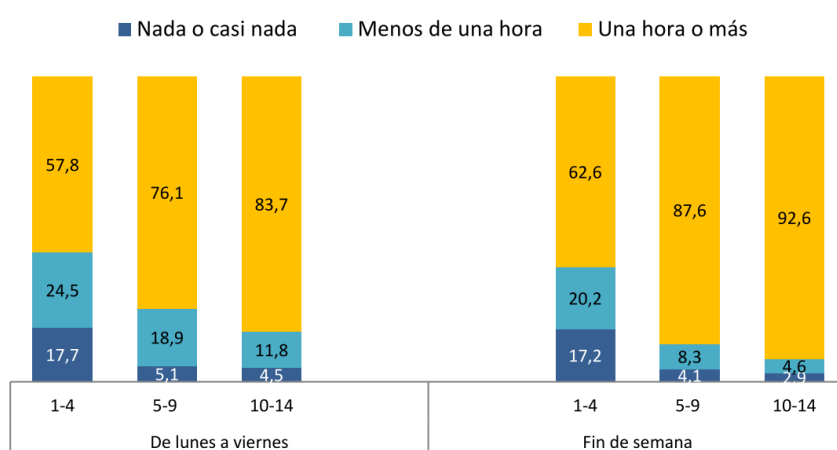


Figura 5. Tiempo libre frente a una pantalla según grupo de edad (%). España 2017. ENSE, MSCBS/INE (35)

En relación al tiempo que los escolares destinan a las tareas del colegio, el estudio ALADINO del año 2019 mostró que el 43,8% emplea como máximo una hora al día entre semana y el 39,9% los fines de semana (Figura 6). Así mismo, describieron que el 24% de los escolares pasa más de 3 horas al día

en actividades de ocio sedentarias, ascendiendo al 39,2% si solo se tiene en cuenta el fin de semana (Figura 7) (49).

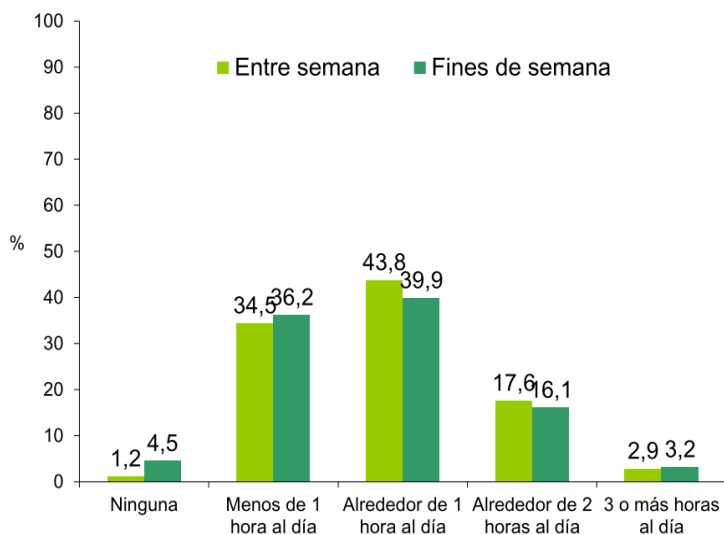


Figura 6. Tiempo dedicado a realizar los deberes o leer libros. España 2019. Estudio ALADINO (49)

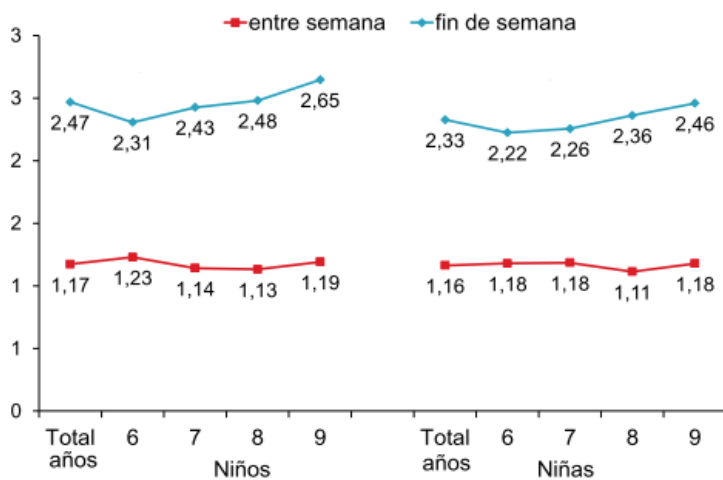


Figura 7. Tiempo libre frente a una pantalla los días entre semana y el fin de semana. España 2019. Estudio ALADINO (49)

Una revisión sistemática del año 2013 demostró que la disposición de equipos electrónicos se asocia positivamente con conductas sedentarias basadas en

pantallas. Aunque los datos fueron insuficientes para determinar si la disponibilidad de material facilitador de AF fomentaba hábitos activos, los investigadores observaron que los niños que disponían de pasatiempos de AF presentaba menos CS. Si bien esta evidencia no es concluyente, plantea la posibilidad de que los juegos activos pueden restar importancia al tiempo frente a una pantalla y disminuir el CS (52). No obstante, algunos autores sugieren que los niños que no disponen de una gran variedad de equipos multimedia pueden, simplemente, sustituir una conducta sedentaria por otra (64).

Las actividades extraescolares representan uno de los principales medios para fomentar hábitos saludables en la infancia. Proporcionan un entorno seguro y estructurado con el potencial para realizar actividades que no se ofrecen durante el horario lectivo o que pueden complementar las materias escolares, incluidos deportes, artes, enriquecimiento cultural, ciencia o educación para la salud (65–67). A diferencia de la investigación sobre el uso de los medios de pantalla y la AF, los estudios sobre las actividades de ocio culturales y artísticas son escasos (41).

Según datos del 2019, el 74% de los escolares españoles están inscritos en actividades deportivas extraescolares, con mayor frecuencia de entre dos (27,8%) a tres (21,7%) horas semanales (Figura 8), siendo el porcentaje superior en los niños y en los grupos de edad de 6, 8 y 9 años (49). Con el incremento de la edad, el porcentaje de población activa en el período extraescolar desciende de manera significativa, de forma que en el tramo final de la etapa escolar dicho porcentaje se sitúa por debajo del 50% (45).

Con respecto al tiempo total dedicado habitualmente a practicar AF de intensidad moderada a vigorosa, los organismos internacionales recomiendan que los niños de entre 5 y 17 años realicen al menos 60 minutos al día de AFMV (3). Se ha descrito que el 78,3% de los niños españoles dedican como mínimo 1 hora diaria entre semana a jugar o hacer ejercicio de forma

intensa, mientras que los fines de semana el porcentaje aumenta al 85% (Figura 9) (49).

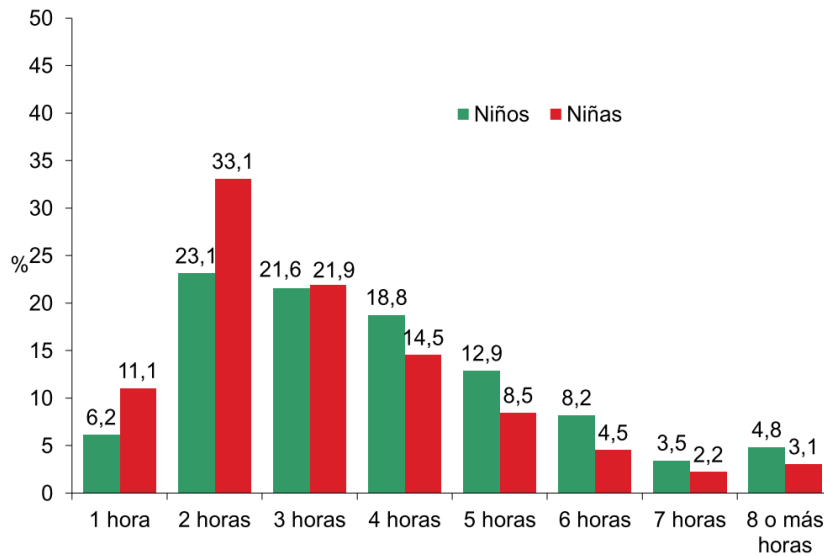


Figura 8. Tiempo libre dedicado a actividades deportivas extraescolares. España 2019. Estudio ALADINO (40)

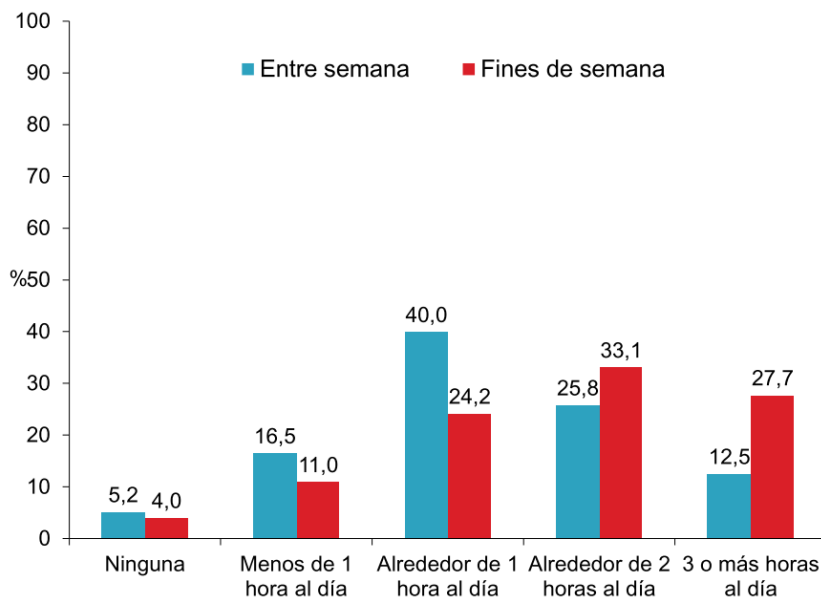


Figura 9. Tiempo libre dedicado a practicar actividad física de intensidad moderada a vigorosa. España 2019. Estudio ALADINO (49)

1.2.2.2. Escuela

La escuela representa un entorno interesante para el desempeño de la AF desde edades tempranas hasta la pubertad. En primer lugar, los niños pasan en la escuela como mínimo 5 horas al día (68). En segundo lugar, disponen de las instalaciones, el material y el personal que pueden no ser accesibles en el hogar o la comunidad (69,70). Los expertos coinciden en que la escuela es un entorno ideal para promover hábitos saludables pero, a pesar de ello, la evidencia demuestra que no está alcanzando su máximo potencial (37). En general, los niveles de AF observados en las escuelas son mucho más bajos de lo recomendado (37,69,71) y una proporción considerable de estudiantes finalizan sus estudios obligatorios sin haber dominado las habilidades básicas de movimiento ni haber obtenido una alfabetización suficiente en AF (70).

Los dos pilares que representan las oportunidades de AF en la escuela son la Educación Física y el tiempo de recreo. El recreo incluye períodos programados regularmente y la mayoría de las escuelas lo ofrecen con mayor frecuencia que la Educación Física, por lo que tiene mayor potencial para contribuir a la AF total (71). La AF también se puede incorporar en el aula de otras formas, como iniciativas de psicomotricidad, yoga u otras propuestas activas que se integren en las lecciones académicas clásicas, como las matemáticas, las lenguas u otras áreas temáticas (72).

Durante el recreo, se han descrito varios factores que pueden condicionar la acumulación de minutos de AF, como el sexo, la edad, la variedad cultural, la estructuración del juego por parte del personal escolar, las condiciones climáticas, la provisión de materiales o el espacio físico del patio (73-75). Por ejemplo, se ha demostrado que las niñas y las poblaciones culturalmente diversas se involucran en comportamientos de intensidad más ligera en comparación con los niños y las poblaciones menos diversas (42,43). Por otro lado, el entorno del patio del recreo puede ser favorecedor de conductas activas y suele ser diferente entre preescolares y niños de Educación Primaria

en lo que se refiere al equipo disponible, a la supervisión de un adulto o a la provisión de espacios (76). Con respecto a la efectividad de diferentes estrategias para aumentar los niveles de AF en el recreo, la evidencia muestra resultados contradictorios (73-75,77,78). Una revisión sistemática del 2020 mostró que cabe la posibilidad de que el tiempo necesario para organizar actividades estructuradas reduzca el tiempo disponible para estar activo durante el recreo y que además pueda minimizar la oportunidad de socializar y aprender habilidades para crear opciones de juego creativas (74).

La Educación Física es la otra fuente importante de AF, pero quizás su papel más importante es proporcionar a los niños el conocimiento, las habilidades y la confianza para ser activos (37,79-81). De hecho, se ha demostrado que las estrategias que incluyen la organización e instrucción de los alumnos combinado con la práctica de actividades de alta intensidad se asocia con un 24% más de aprendizaje activo en comparación con una práctica de Educación Física rutinaria (79).

En España, las leyes educativas marcan las directrices para la organización de la AF en el entorno escolar. La representación de la AF depende principalmente de la combinación de Educación Física y recreo, aunque los centros educativos tienen la posibilidad de implementar innovaciones pedagógicas, programas educativos, formas de organización o ampliación del horario lectivo de ciertas asignaturas, siempre dentro de los términos que establezcan las administraciones y dentro de las posibilidades que permita la normativa aplicable (82).

El estudio ALADINO del 2019 mostró que el tiempo de Educación Física a la semana suele variar de 1 a 3 horas entre centros con una media de 121,1 (30,8) minutos semanales. Aunque el 44,2% de los centros declara invertir 180 min o más a las clases de Educación Física, el 17,4% destina 60 min semanales o menos (Tabla 1) (49). Estas diferencias se deben a que, desde la aprobación de la Ley Orgánica para la Mejora de la Calidad Educativa

(LOMCE) (83) y el Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria (68), son las Comunidades Autónomas y en su caso, los centros escolares, quienes regulan las competencias y horarios de esta asignatura. Con la entrada en vigor de la Ley Orgánica de Modificación de la Ley Orgánica de Educación (LOMLOE), más conocida como Ley Celaá (82), se prevé que los centros educativos sigan ejerciendo su autonomía para regular la Educación Física.

En España, se establece que el recreo dura alrededor de 30 minutos distribuidos a lo largo de la jornada escolar (68). El estudio ALADINO describió que el 97,8% de los centros escolares cuenta con alguna zona de recreo exterior en la que pueden jugar los alumnos y un 77,8% dispone también de gimnasio cubierto (49).

Considerando las recomendaciones de las guías de AF (3), el programa curricular no asegura cubrir los requerimientos de AF en la etapa escolar. La realidad es que los planes docentes actuales contemplan que los niños pasen una gran cantidad de tiempo sentados y existe la tendencia a dedicar progresivamente menos tiempo a los comportamientos activos y aumentar la cantidad de actividades sedentarias (37,69,71).

Tabla 1. Tiempo semanal dedicada a las clases de Educación Física según el nivel escolar. España 2019. Estudio ALADINO (39)

Curso	Min/sem		≤ 60		60-120		120-180		≥ 180	
	Media	DE	n	%	n	%	n	%	n	%
1º EP	124,2	33,7	10	3,6	99	35,8	131	47,6	36	12,9
2º EP	122,3	34,0	12	4,4	99	35,9	133	48,1	32	11,5
3º EP	121,3	33,8	14	5,0	98	35,3	129	46,7	36	12,9
4º EP	116,5	29,7	12	4,5	99	35,9	147	53,3	18	6,4

DE: Desviación estándar, EP: Educación Primaria

1.2.2.3. Transporte

Los desplazamientos representan una opción para complementar los minutos de AF diarios no acumulados en el tiempo libre o en la escuela. La mayoría de los estudios muestran una relación positiva entre los desplazamientos activos y la consecución de niveles significativamente más altos de AF (84,85). Además, reemplazar los viajes motorizados con formas más saludables de transporte también implica numerosos beneficios colaterales, como la reducción de la congestión y accidentes de tráfico o la disminución de los índices de contaminación (86).

Los desplazamientos de ida y vuelta a la escuela son los más importantes a analizar en los niños en etapa escolar. El estudio ALADINO del 2019 mostró que el medio de transporte más frecuente fue caminar (el 51,8% y el 54,6% de los escolares va y vuelve al colegio caminando, respectivamente), seguido del coche o moto (44,6% ida y 42,2% vuelta). En cuanto al uso de transporte público o autobús escolar, se ha descrito que solo un 6% lo utiliza como medio de transporte a la ida y un 6,1% a la vuelta, y en cuanto a la bicicleta o patinete no motorizado el porcentaje es mucho menor (1,7% a la ida y 1,5% a la vuelta). Entre las razones por las que los niños no se desplazan de forma activa a la escuela, destacan la distancia del hogar al centro, seguido de la falta de tiempo y de la inseguridad percibida en la ruta de desplazamiento (49).

1.2.3. Desarrollo de los patrones del movimiento humano

Al igual que los dominios, Pettee Gabriel KK et al. (20) explican los constructos y características generales del comportamiento del movimiento humano en referencia a la población adulta (Figura 2) (20). Sin embargo, el hecho de que los niños presenten patrones de actividad diferentes requiere un enfoque que tenga en cuenta los aspectos únicos del movimiento en la infancia (56).

El desarrollo de los sistemas corporales durante el crecimiento explica la variabilidad en la representación típica de la AF y el CS. Los cambios madurativos en los huesos, músculos y grasa se encuentran entre los más importantes del crecimiento físico ya que alteran las proporciones corporales y definen las diferencias de género y entre individuos (87). Además, se han identificado mecanismos hormonales relacionados con la neurotransmisión de la dopamina durante la locomoción (88) y mediadores hormonales en la pubertad que dan como resultado una reorganización neural y cambios rápidos en el tamaño y la composición corporal (89).

Hay varios períodos clave y puntos de transición de la AF desde el nacimiento hasta la adolescencia. Estos incluyen los primeros años de vida, el inicio de la educación formal (preescolar), la transición de la escuela primaria a la escuela secundaria y la pubertad (90). Se ha descrito que, en general, los patrones de comportamiento siguen un camino hacia la disminución de la AF con la edad cronológica (30).

Los primeros años de vida representan un período intenso de aprendizaje motor que proporciona la base para un desempeño posterior más complejo y hábil (91). Durante los primeros meses, la inmadurez de los sistemas corporales condiciona un comportamiento con predominancia sedentaria, que es saludable y necesario en esta fase del crecimiento. Aún así, se observan patrones de motricidad gruesa difíciles de clasificar, conocidos como “estereotipos rítmicos” (por ejemplo, el balanceo del cuerpo o el movimiento de patadas en las extremidades inferiores). Se cree que la aparición de estos comportamientos probablemente esté controlada por la maduración neuromuscular general y representan manifestaciones de la integración sensoriomotora de los sistemas corporales. Estas estereotipias alcanzan su punto máximo aproximadamente a los 6 meses de vida y después desaparecen gradualmente del repertorio de movimientos. Las interacciones tempranas entre padres e hijos constituyen otra de las oportunidades para la

AF a esta edad, como lanzar al bebé en el aire o rebotar en la rodilla. Así mismo, la manipulación de objetos representa una oportunidad interesante para el movimiento. En los bebés, este comportamiento consiste en ejecutar acciones simples como meter los objetos en la boca o tirarlos al suelo (92).

A medida que progresa el desarrollo, aumentan los períodos de vigilia y se adquieren nuevos hitos motores (gatear, caminar, manipular objetos, etc.), mayor equilibrio, mejor coordinación, etc. Durante el primer año, la AF se caracteriza por la exploración del entorno e implica movimientos repetitivos y centrados en uno mismo (93). A diferencia de los adultos, los niños son intrínsecamente activos porque el movimiento es el medio que les proporciona la estimulación del sistema nervioso central (29). Los niños en esta etapa utilizan el juego como forma para explorar el entorno a través del movimiento y experimentar con las capacidades motrices de su propio cuerpo (56,91). El juego se convierte en el motor del movimiento espontáneo y se interpreta como una elección atractiva en sí misma (94). Típicamente se embarcan en AF transitorias, de corta duración, con grandes descansos y cambios frecuentes de estímulos que no requieran una atención sostenida (95,96).

Al final del primer año de vida aparece el “juego de ejercicio”. Se trata de una actividad dominada por movimientos locomotores gruesos (como correr, saltar, trepar, etc.) cuya característica distintiva es el vigor físico. Se considera que el juego de ejercicio apoya el entrenamiento del sistema músculo-esquelético, especialmente la fuerza y la resistencia muscular (92). Puede ser social o no, pero lo más frecuente es que los niños comiencen con el juego paralelo (uno al lado del otro, pero sin interactuar) y vayan avanzando hasta que se adoptan roles interactivos (97). El juego de ejercicio aumenta desde los niños más pequeños hasta la edad preescolar llegando a su pico máximo alrededor de los 4-5 años (92).

A medida que ingresan en la escuela primaria, se observa una disminución progresiva de la AF. Por un lado, los niños pasan la mayor parte del día en el aula, donde predominan las actividades sedentarias y se dedica poco tiempo a la Educación Física y la participación activa. Con el paso de los años, las actividades escolares ligadas al movimiento van perdiendo peso gradualmente a favor de lecciones puramente académicas, lo que limita las oportunidades de AF y fomenta el tiempo de CS no discrecional (91). Por otro lado, el juego de ejercicio pierde importancia, observándose que de los 6 a los 10 años apenas tiene representación (92). Sin embargo, comienza a desarrollarse la capacidad para participar en actividades grupales más organizadas (97) y aparece un tipo diferente de juego en movimiento conocido como “juego rudo”. El juego rudo se caracteriza por ser un comportamiento vigoroso con actividades como luchar, dar patadas, hacer volteretas, etc. Este tipo de juego involucra ya una conducta social (más frecuente en los varones) y alcanza su punto máximo a los 8-10 años de edad (92).

A medida que continúa el desarrollo de los sistemas corporales, en especial la cognición, el disfrute y la estimulación del sistema nervioso central se encuentran cada vez más en una variedad de actividades no necesariamente locomotoras como la lectura, la expresión artística, la resolución de problemas, las relaciones sociales o las actividades vocacionales (29). El juego activo va perdiendo dominancia como motor del movimiento dejando paso a conductas cada vez más sedentarias (97). En los últimos años de la escuela primaria y la pubertad, la AF observa una caída importante de los niveles espontáneos y se comienza a obtener a través de otros medios de forma consciente y premeditada, como el ejercicio físico planificado o la práctica deportiva (91). Los niños son más propensos que las niñas a mantener la participación en juegos activos, especialmente los juegos de pelota, pero a medida que alcanzan la pubertad, los contextos sociales suelen disuadirlos de

participar en AF y su comportamiento se vuelve más sedentario y se asemeja más al de las niñas (98).

1.2.4. Atributos fisiológicos

Pettee Gabriel KK et al. también definen las consecuencias del movimiento humano que dan como resultado atributos fisiológicos entre los que se incluyen un mayor GE y una mejor aptitud o condición física. Dentro del GE total identifican la tasa metabólica en reposo (TMR), el coste energético asociado con la termogénesis y el GE relacionado con la AF. Además, recogen que el movimiento humano produce mejoras en las dimensiones de la aptitud física relacionada con la salud, incluida la resistencia cardiorrespiratoria, la resistencia muscular, la fuerza muscular, la flexibilidad, el equilibrio, la coordinación y la composición corporal (Figura 2) (20).

1.2.4.1. Gasto energético

La contribución de la TMR y la termogénesis al GE total es relativamente estable. En un individuo promedio, la literatura científica recoge que el consumo asociado al efecto térmico de los alimentos (necesario para la digestión, la absorción y el aumento de la actividad del sistema nervioso simpático posterior a la ingesta) supone aproximadamente un 10%, mientras que la TMR requiere el 60-75% del total. La AF supone el componente más variable y, por lo general, representa un porcentaje relativamente bajo, entre el 15-30% (59,99,100).

La TMR se define como la cantidad de energía necesaria para la integración de la actividad de los tejidos en condiciones de equilibrio cuando el individuo está despierto en un estado térmico neutro y en ayuno. El registro de la TMR requiere un protocolo específico para evitar errores derivados del efecto

térmico de la alimentación, la temperatura ambiental o la práctica previa de AF (101).

La TMR es la base para el equivalente metabólico (MET) que se utiliza habitualmente para cuantificar el GE asociado a una actividad específica. Para un adulto promedio, se considera que corresponde a 1 kcal/kg/h o 3,5 ml O₂/kg/min (102). Se ha descrito que solo varía entre un 5% y un 10% en la vida adulta, y las comparaciones entre grupos de edad, sexo y peso muestran que el 85% de los adultos presentan una TMR que se aleja como máximo un 10% de la media (100). Sin embargo, la TMR por unidad de masa corporal es considerablemente más alta durante la infancia, observándose que disminuye progresivamente con la edad hasta alcanzar los valores normativos aproximadamente a los 15 años en las niñas y a los 16 años en los niños (103,104).

La necesidad de una TMR mayor en la infancia se debe a una variedad de factores, incluidos los costes energéticos adicionales asociados al crecimiento y desarrollo, las diferencias en la distribución de la masa corporal o la etapa puberal (103-105).

Se ha descrito que el metabolismo de los órganos y tejidos, así como su proporción en el cuerpo humano cambian con el desarrollo y condicionan los requerimientos energéticos en reposo. La masa corporal libre de grasa (músculos, corazón, hígado, riñones o sistema nervioso central) constituye la porción de tejidos con mayor tasa metabólica. Con el crecimiento, se observa un incremento de tejido adiposo, de hueso y, en menor medida, del músculo esquelético y el resto de los órganos de alta demanda energética. Debido a que la proporción de masa corporal libre de grasa disminuye proporcionalmente en los adultos a favor de una mayor presencia de masa “energéticamente inerte”, las necesidades metabólicas en reposo son menores con el aumento de la edad (103,105,106). Sin embargo, actualmente se desconoce el detalle de las densidades de los órganos y tejidos en cuanto a su hidratación y

contenido en grasas, por lo que el impacto de la masa de alto requerimiento metabólico aún no se ha definido con exactitud (105).

La TMR también varía según la etapa de la pubertad, el sexo y la etnia. Los requerimientos energéticos asociados a una pubertad precoz o tardía pueden generar diferencias entre individuos de la misma edad, por lo que algunos autores sugieren considerar las etapas de la pubertad de Tanner en lugar de la edad cronológica, especialmente alrededor de los 9-12 años para las niñas y de los 10-14 años para los niños (107). Por último, los estudios muestran que los varones y los individuos de ascendencia europea presentan una TMR ligeramente superior a las mujeres y los sujetos afroamericanos. Este aumento de la TMR se debe probablemente a las diferencias en la distribución y composición corporal y al alcance de la pubertad a una edad más tardía (103,107,108). No obstante, por motivos de practicidad y facilidad de uso, los criterios para clasificar el GE tienden a ignorar la etnia y la etapa puberal a favor del sexo, la edad cronológica y la masa corporal (109,110).

Dada la estabilidad de la TMR y la termogénesis, la medición del GE relacionado con la AF se suele utilizar para inferir el consumo total (100,111). El GE relacionado con una actividad refleja el costo asociado a los procesos metabólicos involucrados en el intercambio de energía requerido para la contracción del músculo esquelético (111). Siendo estrictos, el registro del GE durante una AF corresponde al coste bruto, por lo que para calcular el coste neto de la actividad en sí misma se tendría que restar al valor total la TMR y el efecto térmico de la alimentación. Para las actividades de una intensidad alta en las que la TMR representa un bajo porcentaje del total (un 10% o menos) tiene poca relevancia ajustar el GE para calcular el costo neto. Sin embargo, es importante considerarlo en las actividades de intensidad ligera (59).

Dentro del GE asociado a la AF se incluye la postura, la AF espontánea y la AF voluntaria (100). La consideración de la postura como parte de la AF ha sido

objeto de debate ya que implica poco o ningún movimiento. Sin embargo, se ha admitido que se diferencia de los CS en la posición de verticalidad y en que requiere una contracción isométrica de los músculos anti gravitatorios (21). Por otro lado, la AF espontánea se refiere a las actividades relacionadas con el desempeño de la vida diaria. Se ha descrito que es el principal determinante del GE asociado a la AF y supone un reservorio cuidadosamente regulado por principios biológicos, entre ellos, factores neurales, periféricos y humorales (29,112,113). Tal y como se ha explicado en el capítulo 1.2.3, la AF espontánea tiene un rol importante en el comportamiento del movimiento en la infancia. Los niños son intrínsecamente activos porque el movimiento les proporciona la estimulación del sistema nervioso central, lo cual indica que sus patrones de AF espontánea se presentan con gran frecuencia (29,56,96).

Independientemente de si la AF se ha planificado o si se realiza como parte de las rutinas diarias, es fundamental caracterizarla en función de los costos metabólicos que implica. La forma más precisa de registrar el GE es a través de métodos directos, como la calorimetría o el agua doblemente marcada (capítulo 341.3.1). Sin embargo, debido a los costes asociados a estos procedimientos, no siempre es posible calcularlo de forma directa. Una buena alternativa para estimar el GE es caracterizar el patrón de actividad combinando las variables de frecuencia, duración e intensidad (56,100). El producto de estas tres variables resulta en el GE total durante un tiempo específico y se toma como una medida del volumen de AF. El volumen se puede describir en valores absolutos (por ejemplo, kilocalorías) pero habitualmente se expresa como MET-min o MET-hora (59).

La frecuencia se puede definir como el número de veces que se realiza una determinada actividad dentro de un rango de intensidad específico y dentro de un período de tiempo predeterminado, mientras que la duración se refiere a la cantidad de tiempo que perdura la actividad (20,59). Aunque ambos representan un concepto sencillo, su registro es bastante complejo, ya que los

niños se involucran constantemente actividades intermitentes de corta duración que contribuyen al volumen acumulado de AF en el mismo grado que una actividad de la misma intensidad realizada en una sesión más larga. Esta es una modificación recientemente incluida en la guía de AF de la *American College of Sports Medicine (ACSM)*, que corrige la indicación de la guía anterior que recomendaba no contabilizar los períodos inferiores a 10 minutos (3).

El concepto de intensidad se define como el nivel de esfuerzo o demanda fisiológica necesario para realizar una actividad específica (20). Se calcula como el GE por unidad de tiempo y habitualmente se expresa como el consumo de O₂ por unidad de masa corporal o como un múltiplo de la TMR expresada como MET (100,102). El cálculo de la intensidad de una actividad en METs se basa en el principio de eficiencia mecánica. Si se supone una eficiencia constante para realizar una actividad, se puede deducir que el GE neto calculado en un entorno controlado de laboratorio será muy similar al GE real en la vida diaria (102). Habitualmente se divide en cuatro categorías: intensidad vigorosa (≥ 6 METs), intensidad moderada (3 - 5,9 METs), intensidad ligera (1,6-2,9 METs) y CS ($\leq 1,5$ METs). Todas las actividades que requieren un GE igual o menor a 1,5 METs encajan dentro de la definición de CS ya que implican que el individuo esté sentado, reclinado o acostado. La única excepción es el mantenimiento de la bipedestación, que como ya se ha explicado, se considera AF aunque su GE esté dentro de los límites del CS (3).

La clasificación de las AF según su intensidad, aunque no tiene en cuenta las particularidades individuales, proporciona un sistema de clasificación muy útil que estandariza las medidas de AF autoinformadas (100,102). En 1993, Ainsworth et al. desarrollaron el primer compendio de AF y CS (102). Fue revisado en el 2000 (114) y 2011 (115) y actualmente es considerado una referencia para los valores de intensidad en adultos sin limitaciones físicas, pero se desaconseja su uso en niños (102,116).

Una las principales razones por las que se debe evitar el uso del compendio de adultos es la relación entre el MET y la TMR. Al expresar la intensidad en METs, se asume que se está ofreciendo un valor múltiplo de la TMR (100,102). Por tanto, si la TMR es mayor en niños, en general, el GE durante una AF aumenta menos proporcionalmente, por lo que utilizar el valor de 3,5 ml O₂/kg/min no es adecuado para representar 1 MET (104,116).

Por otro lado, el GE de un niño está condicionado por diferencias en la eficiencia mecánica. En primer lugar, los niños son poco eficientes al realizar muchas actividades debido a los factores ambientales propios de un entorno adulto (por ejemplo, la altura de una mesa, el peso de un libro, el tamaño de una taza, etc.) y su falta de experiencia (104,117,118). La influencia del entorno y la experiencia pierden importancia en actividades como correr o caminar, ya que los niños pueden seleccionar una velocidad y patrón propios sin tener que adaptarse a accesorios de tamaño adulto. Además, son actividades muy practicadas en las que el niño con cierta edad tiene suficiente bagaje como para haber desarrollado una eficiencia superior a otras actividades menos usuales, por ejemplo, las tareas del hogar (118). Por otro lado, se ha de tener en cuenta que las habilidades para ejecutar diferentes actividades se adquieren de forma no homogénea, por lo que se desaconseja el uso de un único valor estándar para toda la franja de pediatría, ya que sobrestimaría el GE en los niños más pequeños y lo subestimaría en los adolescentes (117,119).

Aunque se ha avanzado mucho en el desarrollo y actualización del compendio para adultos, ha habido menos investigación sobre el GE asociado a la AF y CS en la infancia (104,117,118,120–122). Las primeras propuestas para estandarizar el GE en niños sugerían ajustar el compendio de Ainsworth et al. según la TMR por edad o etapa puberal (104), aunque no había consenso sobre la mejor forma de hacer las estimaciones (120). En 2008 se publicó un primer compendio con datos referentes a la franja de pediatría identificados

a través de una revisión de la literatura científica (118). Sin embargo, la falta de información condicionó que solo el 35% de los valores se tomaran de estudios empíricos, mientras que los datos restantes se extrapolaron del compendio de adultos (114).

Para abordar estas limitaciones y actualizar el compendio juvenil se convocó el *Youth Energy Expenditure Workshop* (123). El panel de expertos acordó revisar la literatura para identificar estudios no recogidos en trabajos anteriores (118), discutir la métrica más adecuada (124) y solicitar datos de estudios no publicados (121). Seis años después se publicó un compendio actualizado, amplio y disponible que presenta el GE de 196 actividades derivadas únicamente de datos pediátricos, expresados en términos MET ajustados por rangos de edad (6–9, 10–12, 13–15 y 16–18 años) (117). Sin embargo, aún quedan áreas críticas que requieren un estudio en mayor profundidad. Entre ellas, el estudio del GE de las actividades en niños menores de 6 años, la ampliación del compendio para incluir actividades más específicas de transporte o ciertos deportes y la adaptación del compendio ya existente en poblaciones con limitaciones físicas que alteren la eficiencia mecánica durante el desempeño de una actividad (116).

1.2.4.2. Efectos de la actividad física y el comportamiento sedentario

Para producir una estimación no sesgada de la fuerza de asociación entre el comportamiento del movimiento humano y un resultado de salud específico se debe determinar si existen pruebas que sustenten una evidencia adecuada en apoyo de una relación causal (111). La dosis-respuesta se refiere a esta relación entre los niveles crecientes de AF o CS y los cambios en un parámetro de salud definido (59).

Se deben considerar varios aspectos en la evaluación de una posible relación dosis-respuesta. En primer lugar, la idea de que toda AF produce beneficios

para la salud está abierta a crítica, ya que actividades que impliquen esfuerzos repetitivos no producen beneficios para la salud y, sin embargo, se contabilizan como AF (19). En otras palabras, para poder establecer una relación de causalidad entre la práctica de AF y la mejora del estado de salud la proporción dosis-respuesta debe ser adecuada y no debe exceder la capacidad del individuo (59,111). En segundo lugar, los comportamientos de movimiento son solo uno de los muchos factores que pueden influir en los indicadores de salud en la infancia y es complejo aislar los efectos atribuidos a la AF o el CS de los asociados con la nutrición, el crecimiento, la maduración y el desarrollo normales (1,3).

Con base en la ciencia existente, no es posible ser muy preciso en la selección de una sola expresión de volumen de AF debido a la diversidad en los tipos de actividades, las condiciones en las que se realizan, los diferentes instrumentos, las distintas unidades de medida y las particularidades en la relación dosis-respuesta en función del resultado de salud. Específicamente, las pautas para niños están diseñadas para garantizar un crecimiento y desarrollo saludables, minimizar los riesgos de una mala adaptación a la sobrecarga, prevenir lesiones y brindar oportunidades para participar de manera placentera en una amplia gama de actividades. En este contexto, la evidencia científica sugiere que los beneficios para la salud se consiguen con la acumulación diaria de 60 minutos o más de AFMV. Además, se especifica que deben incluirse ciertos tipos de AF durante 3 o más días a la semana, como actividades de fuerza y resistencia muscular, actividades de predominancia aeróbica y actividades de carga de peso (1,3).

En niños de 3 a 5 años existe evidencia sólida de que el volumen acumulado de AF se asocia con un riesgo reducido de aumento de peso excesivo y adiposidad (125–127) y con indicadores favorables de salud ósea (128–130), pero no se dispone de evidencia suficiente para determinar los efectos de la AF sobre los factores de riesgo cardiovasculares y metabólicos (131,132).

Para los niños en edad escolar y los adolescentes, se ha demostrado que la AF regular se asocia con una mejor resistencia cardiorrespiratoria (84,133-136), una mayor resistencia y fuerza muscular (137-140), una disminución del peso y porcentaje de grasa corporal (141-143), una mejor salud ósea (142,144-146), un menor riesgo de enfermedades cardiovasculares y metabólicas (136,147,148), una reducción de los síntomas de depresión o ansiedad (149-152) y una mejora en los indicadores de la cognición, entre ellos, la memoria, la velocidad de procesamiento, la atención y el rendimiento académico (153-155).

Con respecto al impacto de las CS sobre el estado de salud en niños y adolescentes, la evidencia sugiere que una mayor proporción de sedentarismo está relacionada con una peor salud cardiovascular y metabólica y un mayor índice de adiposidad en niños y adolescentes (156-158), pero no se ha podido demostrar que esté relacionada con un empeoramiento de la salud ósea (159,160). Al igual que ocurre con las recomendaciones específicas del volumen de AF, no se dispone de evidencia suficiente para determinar la relación dosis-respuesta entre un mayor tiempo dedicado a los CS y peores resultados de salud (1,3).

1.3. Métodos de medición de la actividad física y el gasto energético

La medición del comportamiento de movimiento humano no solo es necesaria para identificar asociaciones causales entre la práctica regular de AF y los resultados de salud o avanzar en el estudio de las relaciones dosis-respuesta comentadas en el capítulo anterior, sino que, además, es imprescindible en la formulación de recomendaciones de salud pública, en la validación de programas de intervención o en la comparación de los niveles de actividad intra e inter individuos (100).

Para comprender la amplitud de los métodos de evaluación de la AF nos referimos de nuevo al marco conceptual propuesto por Pettee Gabriel KK et al. (Figura 2) (20). A pesar de que muchos autores utilizan indistintamente medidas de AF y de GE, es importante recordar que la AF es la manifestación del comportamiento humano y el GE es el coste derivado de ese comportamiento (20,111). Así, los métodos para medir el GE son aquellos basados en enfoques fisiológicos que registran de forma directa o indirecta el coste derivado de la AF, como la calorimetría, el agua doblemente marcada, la frecuencia cardíaca (FC), la ventilación o la temperatura corporal. Al entender la AF como el reflejo del comportamiento, los métodos de medición adecuados son aquellos que recogen el movimiento ya sea de forma directa, como los acelerómetros o los podómetros, o de forma indirecta, como los diarios o los cuestionarios de AF. Estos métodos permiten capturar la esencia del comportamiento, pero también pueden llegar a estimar el GE derivado de las AF registradas (111).

Siguiendo la propuesta de estos autores, el propósito de este capítulo es exponer las características de los instrumentos de medición de AF disponibles clasificados según el constructo que pretenden medir (Figura 10).

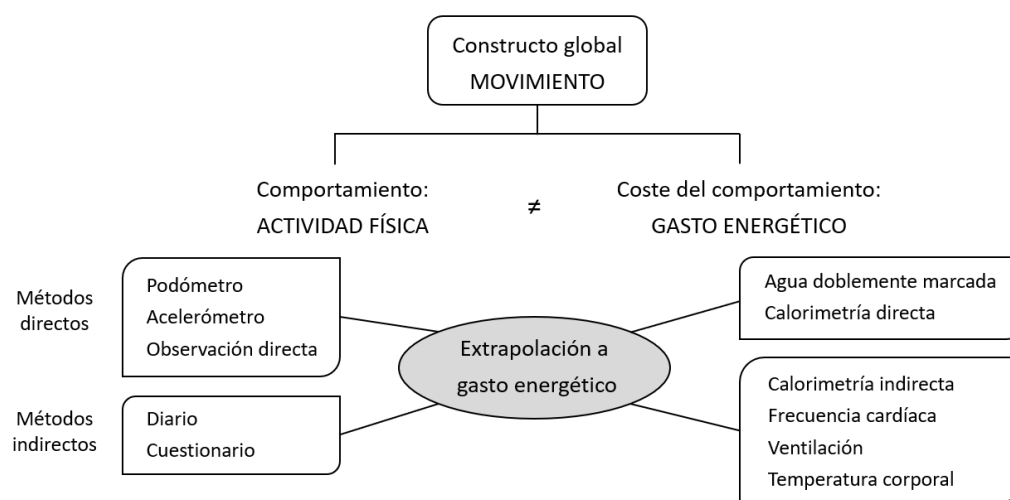


Figura 10. Modelo conceptual de los métodos de medición de la actividad física y el gasto energético. Adaptación de Lamonte et al. 2001 (111)

1.3.1. Instrumentos de medición del gasto energético

Los métodos directos para medir el GE son el agua doblemente marcada y la calorimetría directa. Los métodos indirectos estiman el GE a partir de otras medidas fisiológicas como el consumo de oxígeno (VO_2), la FC, la ventilación o la temperatura.

El agua doblemente marcada se utilizó por primera vez en humanos a principios de la década de los 80 (161–163) y se validó su uso en niños en 1987 (164). Se trata de un método en el que el sujeto ingiere los isótopos estables y no radiactivos 2H_2O y $H_2^{18}O$ que se distribuyen por todo el cuerpo y se excretan gradualmente en la orina. La velocidad a la que se eliminan los isótopos del organismo es proporcional al grado de producción metabólica de CO_2 (VCO_2). Por tanto, el VO_2 y, en consecuencia, el GE total se pueden calcular para el período de estudio a partir de la diferencia en las tasas de eliminación de los isótopos usando ecuaciones establecidas. Dependiendo de la dosis administrada y de la tasa de excreción, el período de recolección de orina habitualmente se extiende de 1 a 2 semanas (58,100,111,165,166).

El agua doblemente marcada es particularmente útil para la evaluación en condiciones de vida libre, ya que no requiere monitores ni requisitos especiales de laboratorio durante el período de medición (58,100,111). Es un método seguro, preciso, no invasivo y puede utilizarse en niños (56). Sin embargo, los isótopos son inasequibles para la mayoría de los investigadores y el procedimiento de recolección es muy exigente, ya que requiere un gran compromiso para recoger todas las muestras (15). Así mismo, se ha descrito una fuente de sesgo que altera el comportamiento del sujeto en estudio debido a la conciencia de ser observado (efecto Hawthorne), por lo que se ha cuestionado la representatividad real de los niveles de AF evaluados con este método (167). Además, la muestra de padres y niños que están dispuestos a recolectar muestras de orina durante semanas, puede diferir bastante de la población de estudio típica (15). Por tanto, actualmente su uso está

desplazado a la validación de otros métodos más prácticos y asequibles (100,111,168).

La calorimetría directa es otro método de medición del GE que se basa en el principio de que toda la energía se transforma en trabajo o calor. El procedimiento de medición se realiza en cámaras herméticas con paredes aislantes que miden la temperatura y estiman la producción de calor mediante un proceso técnico muy complejo. Se trata de un método costoso, poco práctico y limitado a tareas específicas dentro del laboratorio, por lo que actualmente está casi en desuso (100,168).

En sustitución de la calorimetría directa, se ha implementado un método de medición indirecto que mide el VO_2 y se correlaciona estrechamente con la producción de calor. El procedimiento se realiza también en el entorno de un laboratorio y requiere que el sujeto de estudio lleve una mascarilla con un recipiente acoplado para la recolección del aire espirado. La forma más utilizada de calorimetría indirecta emplea un sistema de circuito abierto en el que se respira una mezcla de gases de concentración conocida que permite analizar el VO_2 y el VCO_2 (169–171). A día de hoy, existen métodos de calorimetría indirecta transportables capaces de registrar el GE fuera del entorno del laboratorio. Sin embargo, son engorrosos, caros y afectan a la AF habitual del individuo por el efecto Hawthorne (100,167). Por tanto, el uso de la calorimetría indirecta ha quedado desplazado a estudios de laboratorio o estudios de validación de otros instrumentos (58,100,111,172).

La monitorización de la FC es otra medida indirecta que se deriva de la conexión fisiológica entre el incremento en los latidos del corazón y el estrés cardiorrespiratorio durante la AF. La estimación del GE a través de este método asume que existe una estrecha relación lineal entre la FC y el VO_2 (58,168,173). Sin embargo, se presentan varios problemas. En primer lugar, la relación lineal parece que solo ocurre a niveles de intensidad alta y es poco precisa para evaluar el GE a bajas intensidades (56,100,173,174). En segundo

lugar, tanto al inicio como al final de una AF hay un período de adaptación de la FC que no corresponde con el VO_2 y que puede sesgar la actividad esporádica de los niños o sobreestimar la duración durante el tiempo de recuperación (58,95). Por último, la relación entre la FC y el VO_2 se ve afectada por el volumen de masa muscular involucrado en la actividad, el tipo de contracción muscular, la temperatura ambiental, la aptitud física del individuo, el nivel de fatiga, el estado emocional o determinados fármacos (56,58,111,172).

La viabilidad de este método se ha incrementado con el desarrollo de receptores de muñeca asequibles que pueden almacenar datos de alta resolución (58). Sin embargo, para establecer los parámetros individuales, se necesita previamente una prueba de esfuerzo en laboratorio que permita desarrollar gráficos que expresen la relación entre la FC y el VO_2 . Esta evaluación es costosa y poco práctica para la mayoría de las investigaciones (58). En su lugar, se han desarrollado enfoques generalizados que estiman la FC máxima mediante ecuaciones predictivas, como la ecuación de Fox o la de Tanaka. Una revisión sistemática del 2019 mostró que en niños y adolescentes la ecuación de Fox sobreestima la FC máxima en 12,4 (16,2) lpm, mientras que la ecuación de Tanaka la subestima en 2,7 (5,8) lpm. Estos hallazgos muestran que las predicciones de la FC máxima derivadas de poblaciones adultas no son aplicables en niños (175). Por todos estos motivos, los monitores de FC raramente se utilizan como medida del GE en estudios observacionales, pero están muy extendidos en estudios experimentales. Los estudios de intervención suelen utilizar un porcentaje de la FC máxima real o estimada para prescribir un programa de AF y garantizar que el nivel de esfuerzo sea apropiado para los participantes (3,100).

El registro de la ventilación durante la AF también podría estimar el GE de forma indirecta, pero todavía no se ha desarrollado ningún dispositivo que

cumpla los requisitos mínimos en cuanto a propiedades psicométricas, viabilidad y practicidad (100,111,176).

Por último, la medición de la temperatura corporal se ha descrito como un posible método en la estimación del GE pero se desaconseja su uso porque se altera en gran medida por las condiciones ambientales y por factores individuales (111).

1.3.2. Instrumentos de medición de la actividad física

Los métodos directos para medir la AF incluyen los monitores que registran el movimiento o la aceleración del cuerpo, como los podómetros o los acelerómetros y la observación directa del sujeto en estudio. Los métodos indirectos se basan en la información proporcionada por el individuo para registrar actividades a medida que ocurren (diarios) o para recordar actividades anteriores (cuestionarios).

Los podómetros son dispositivos que se utilizan para contabilizar los pasos durante los ciclos de la marcha en un período determinado de tiempo (58,177,178). Los primeros modelos estaban basados en un engranaje mecánico con resortes y fueron fabricados con la intención de medir la distancia total recorrida, aunque con grandes errores de precisión. El diseño evolucionó a un sistema de brazo de palanca pendular horizontal con un resorte que se mueve con las aceleraciones y registra un paso cada vez que entra en contacto con un sensor. Actualmente, existen podómetros electrónicos basados en mecanismos piezoeléctricos que han mostrado una mayor precisión en comparación con sus predecesores, especialmente para la locomoción a baja velocidad (174,179,180).

Los podómetros están diseñados para registrar los movimientos en un único eje, lo que los convierte en dispositivos insensibles a las AF no locomotoras.

Por ejemplo, el movimiento de las extremidades superiores, el transporte de una carga de peso, los desplazamientos por planos inclinados, los saltos, las volteretas, etc. no se registran adecuadamente con este método (56,95,174,177). En consecuencia, los podómetros proporcionan una estimación del volumen de AF muy sesgada en la que se asume que la mayor parte de la actividad implica un movimiento locomotor (58). Además, solo son capaces de registrar el recuento de pasos totales sin ofrecer información sobre la frecuencia, la duración o la intensidad de la actividad. Así mismo, se ha descrito un posible cambio de comportamiento derivado del efecto Hawthorne que cuestiona la idoneidad del método (167). En general, se necesita más investigación para demostrar que son instrumentos válidos y fiables para evaluar los patrones de AF en niños (181,182).

Los acelerómetros son los sensores de movimiento más comúnmente utilizados para registrar la AF. Utilizan transductores piezoeléctricos y microprocesadores para cuantificar la magnitud y la dirección de la aceleración producida por la acción de la musculatura esquelética en uno, dos o tres ejes (100,177,183). Se colocan habitualmente en la cintura, pero también pueden ir acoplados a la muñeca, al tobillo o al muslo (111,174,177).

Los acelerómetros son sensibles a los patrones de actividad intermitente característicos de los niños en condiciones de alta y baja intensidad (56,95), siempre y cuando se seleccione un procedimiento de análisis adecuado que permita detectar la actividad esporádica (9). A diferencia de los podómetros, los modelos triaxiales sí son capaces de registrar todos los planos y cuantificar con mayor precisión los movimientos producidos por los niños durante el juego, como rodar, trepar, saltar o dar volteretas (56,95). Sin embargo, algunas actividades no implican variaciones en la aceleración, como la contracción muscular isométrica, o requieren un mayor esfuerzo del que es capaz de registrar, como el ejercicio de fuerza contra resistencia, el transporte de una carga de peso, etc. Por tanto, es probable que los acelerómetros

subestimen el volumen total de AF si se realizan frecuentemente este tipo de actividades (95,100). Por otro lado, a día de hoy siguen teniendo un coste elevado y se necesita personal cualificado para descargar e interpretar los datos (178). El análisis de los datos brutos de salida permite obtener información sobre la frecuencia, la duración y la intensidad para clasificar la AF en diferentes niveles a través de ecuaciones de regresión. Sin embargo, no existe consenso sobre los puntos de corte apropiados y los criterios utilizados tampoco están libres de error (174,184,185). Además, se atribuye un posible efecto Hawthorne asociado al comportamiento de la AF durante los días de medición (167).

Para finalizar con la revisión de los métodos directos, algunos investigadores consideran que las técnicas de observación son muy interesantes en los estudios con niños (56). Si bien la información proporcionada por este método puede ser muy útil para caracterizar la AF, no es práctica para desarrollar estudios observacionales de más de unas horas. En primer lugar, se requiere mucha implicación por parte de los investigadores para registrar todas las actividades que realiza un niño. En segundo lugar, es probable que solo grupos seleccionados consientan ser observados continuamente. Por último, se ha descrito que las propias observaciones alteran el comportamiento típico haciendo que no pueda ser considerado como representativo de lo habitual (100,168).

En los métodos indirectos, el registro de los patrones de actividad generalmente proviene de un autoinforme sobre el propio comportamiento. No obstante, en poblaciones especiales, como las personas con discapacidad intelectual, los ancianos o los niños pequeños, la información puede provenir de interlocutores indirectos. En los niños pequeños, es frecuente que los informadores sean los padres o los profesores, dependiendo del contexto (15,178).

Los diarios de AF se basan en solicitar un registro de los diferentes tipos de actividades realizadas y el tiempo dedicado a cada una de ellas durante un período de tiempo específico (100). Por lo general, contienen una lista de actividades que facilita la recogida de información y el procesamiento de los datos según los costes MET asociados a cada actividad (100,111,178).

La principal ventaja de los diarios es su exhaustividad en la información recogida sobre el tipo, la duración, la frecuencia y la intensidad de la AF. Sin embargo, se ha descrito que el proceso de registro constante y el deseo de aceptabilidad social pueden producir cambios en los patrones de comportamiento o llevar al respondedor a exagerar sus respuestas (100,111,167). Además, la técnica del diario presenta grandes impedimentos logísticos debido a las altas exigencias que se imponen en su cumplimiento, especialmente en población pediátrica (15).

Por último, los cuestionarios se utilizan para identificar las dimensiones y los dominios de los comportamientos de AF a partir de respuestas autoinformadas o proporcionadas por un cuidador. Los cuestionarios varían en su detalle, desde unos pocos ítems hasta un historial extenso durante el último año o incluso durante toda la vida. Según el grado de detalle, los cuestionarios de AF se clasifican en tres categorías: cuestionarios de historial global, cuestionarios de recuerdo y cuestionarios de historia cuantitativa (58).

Los cuestionarios de historia global son encuestas breves de uno a cuatro ítems que proporcionan una descripción general del nivel de AF. Este tipo de cuestionarios carecen de la capacidad de capturar todas las actividades diarias, pero dan una estimación rápida de algunos componentes (100). Habitualmente están diseñados para ofrecer una clasificación simple del individuo, por ejemplo activo o inactivo, pero facilitan información muy limitada sobre los tipos y patrones específicos de AF (111). La principal ventaja de los cuestionarios globales es su simplicidad y facilidad de

administración y la principal desventaja es la dificultad para medir el cumplimiento de las directrices de AF y para establecer relaciones dosis-respuesta (58,174).

Los cuestionarios de recuerdo requieren que los encuestados identifiquen la frecuencia, duración e intensidad de tipos específicos de AF realizados en el último día, semana o mes. Habitualmente contienen de 7 a 20 ítems y pueden ser auto administrados, administrados por un entrevistador o cubiertos por un cuidador (58,111,174). En comparación con los cuestionarios de historia global, son un poco más complejos y difíciles de completar, pero la evaluación de AF es más detallada (111). La principal ventaja de los cuestionarios de recuerdo es la capacidad de medir el cumplimiento de las pautas de AF y de establecer relaciones dosis-respuesta con los resultados del estudio. La principal desventaja es la dificultad para recordar los detalles de la AF y la incapacidad de algunos encuestados para promediar las frecuencias y duraciones durante períodos pasados (174).

Los cuestionarios de historia cuantitativa son encuestas detalladas que preguntan por las actividades realizadas durante el último mes, el último año o toda la vida. Habitualmente incluyen entre 20 y 60 preguntas. Debido a su extensión y los requisitos de memoria a largo plazo, suelen ser administrados por un entrevistador (58,100,111,174). Se utilizan generalmente en estudios epidemiológicos para estimar el impacto a largo plazo de la exposición a AF sobre los resultados de morbilidad y establecer relaciones dosis-respuesta. La principal desventaja es la dificultad para recordar los detalles de la AF en un período de recuerdo tan amplio (58,174).

En comparación con los métodos directos, los cuestionarios requieren poco coste económico y de tiempo, son fáciles de distribuir y administrar, no requieren mucha implicación y permiten recopilar información en un gran número de sujetos, lo que los convierte en el método de elección en los estudios epidemiológicos (58,100). Otro punto fuerte es que son capaces de

identificar las dimensiones y los dominios en los que se desarrolla la AF (172,177). Además, no tienen riesgo de afectar a los patrones habituales porque el período de referencia para medir la AF es anterior a su administración (111,186).

Sin embargo, existen ciertas limitaciones generales con respecto a su uso. En primer lugar, el sesgo de deseabilidad social puede condicionar una mala interpretación deliberada de las conductas habituales y puede llevar a sobreestimar la duración de las AF e infraestimar la duración de los CS (186,187). En segundo lugar, las demandas cognitivas subyacentes al almacenamiento de recuerdos (188) junto con los modelos que explican su recuperación (189) son procesos complejos y sujetos a error, especialmente en niños. En consecuencia, las respuestas de los cuestionarios dependen del sesgo de memoria a la hora de definir con exactitud los patrones de AF (111,172,178). Además, la naturaleza esporádica de la AF de los niños hace que estas actividades sean aún más difíciles de recordar, cuantificar y categorizar en términos de frecuencia, duración e intensidad (15).

1.3.3. ¿Cómo escoger un instrumento de medición?

Aunque algunos investigadores consideran que el agua doblemente marcada (165), la calorimetría (170) o la observación directa (15) son los gold standard en la medición de los niveles de AF, muchos otros defienden que no existe un gold standard real (20,111,177,190–192). La evidencia científica sugiere que ningún método puede capturar todos los subcomponentes de la AF y la selección de uno u otro se debe basar en criterios objetivos en función de las indicaciones de uso, la practicidad, las propiedades psicométricas y el objetivo de investigación (20,58,172,174).

Según Lagerros et al., si la medida de exposición clave es el GE, se requieren métodos basados en enfoques fisiológicos, pero el costo y las molestias

asociadas limitan su viabilidad. Estos autores consideran que los instrumentos de autoinforme que identifican la frecuencia, duración e intensidad de la AF probablemente son los más adecuados para analizar las asociaciones entre la AF y los indicadores de salud, aunque a día de hoy presentan importantes errores en la estimación de la AF. En general, sugieren que la elección de cualquier instrumento debe ser un equilibrio entre la validez, la fiabilidad y la viabilidad, entendida esta última como la aceptabilidad del instrumento por parte del usuario, los costes económicos y la carga al participante (100). No obstante, centrar únicamente la elección en estas tres variables puede obviar algunos atributos u otras propiedades de medición relevantes para ciertos fines (por ejemplo, la capacidad de respuesta en estudios evaluativos) (192).

En el 2010, la Asociación Europea de Prevención y Rehabilitación Cardiovascular estableció una guía para facilitar la elección de un instrumento de medición de AF (Figura 11). En ella se consideraba el objetivo de la evaluación como principal determinante, dentro del cual se incluían, además, el tipo de AF, el GE asociado, el tiempo empleado en diferentes intensidades, la duración, la frecuencia, la discriminación por niveles de actividad o las relaciones dosis-respuesta entre la AF y resultados de salud específicos (172). Dos años después, se propuso el modelo conceptual del comportamiento del movimiento humano en el que también se proponían estrategias para mejorar la selección de herramientas de evaluación de la AF (Figura 12). Se postularon cuatro puntos principales: las características del estudio, las características de la población, las características de cada instrumento y las características de los patrones de actividad (20). En 2013, la Asociación Americana del Corazón presentó una matriz de decisión más completa para guiar la selección del instrumento de AF que mejor se adapta al objetivo de evaluación y al entorno (Figura 13). El proceso incluía identificar la variable principal del estudio y valorar aspectos relacionados con la viabilidad y practicidad de la herramienta, como el tamaño de la

muestra, la carga para los participantes, los costos de tiempo, la disponibilidad de recursos y personal, los requisitos de procesamiento y análisis de datos o la necesidad de retroalimentación inmediata (58).

En particular, parece que los cuestionarios están indicados cuando se quiere conocer el tipo de actividad practicada, el tiempo dedicado a diferentes tareas o el detalle de las dimensiones de la AF. Se recomienda su uso con fines discriminativos, para evaluar posibles relaciones dosis-respuesta o para verificar si un individuo cumple con las recomendaciones de las guías de AF. Además, son los instrumentos de elección cuando se dispone de pocos recursos económicos o de personal, cuando se pretende evaluar un número elevado de sujetos, cuando se quiere reducir la carga al paciente o cuando se necesita una rápida interpretación de los datos (58,172).

En resumen, los cuestionarios no deben infravalorarse con respecto a los métodos directos, pues son instrumentos con entidad por derecho propio y en muchas ocasiones son además la elección óptima (193).

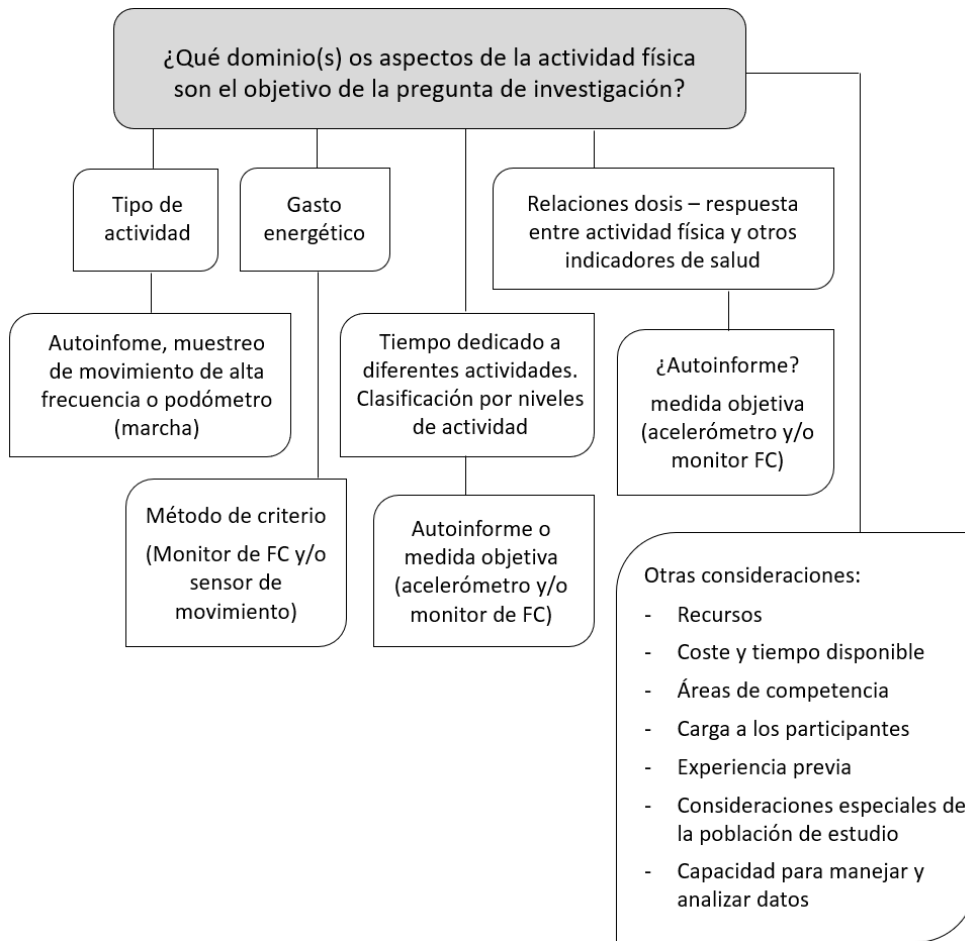


Figura 11. Algoritmo de decisión. Adaptación de The European Association of Cardiovascular Prevention and Rehabilitation. 2010 (172)

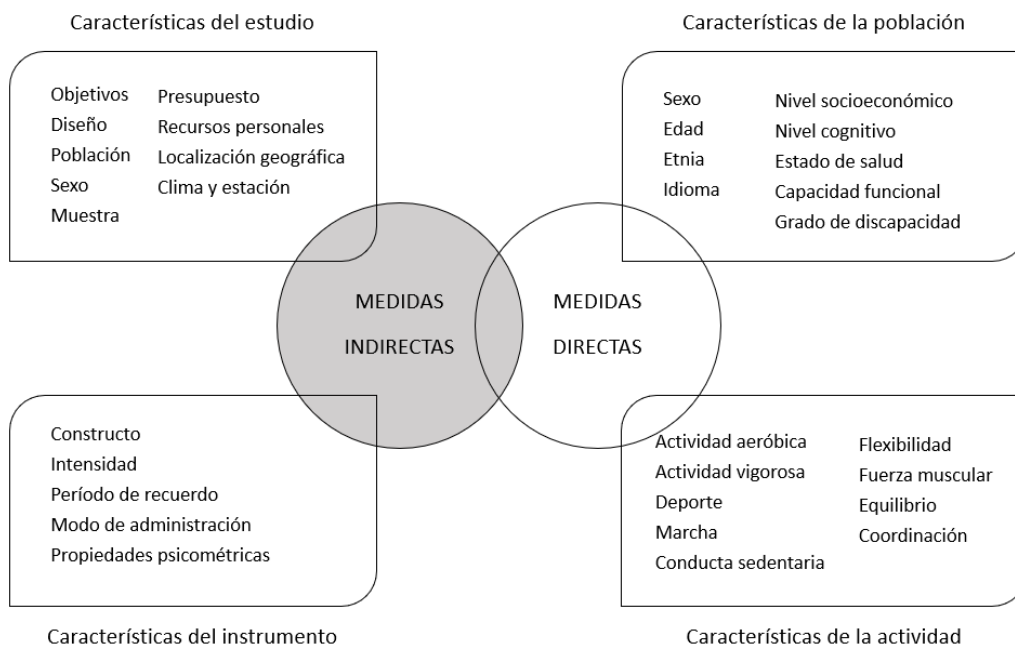
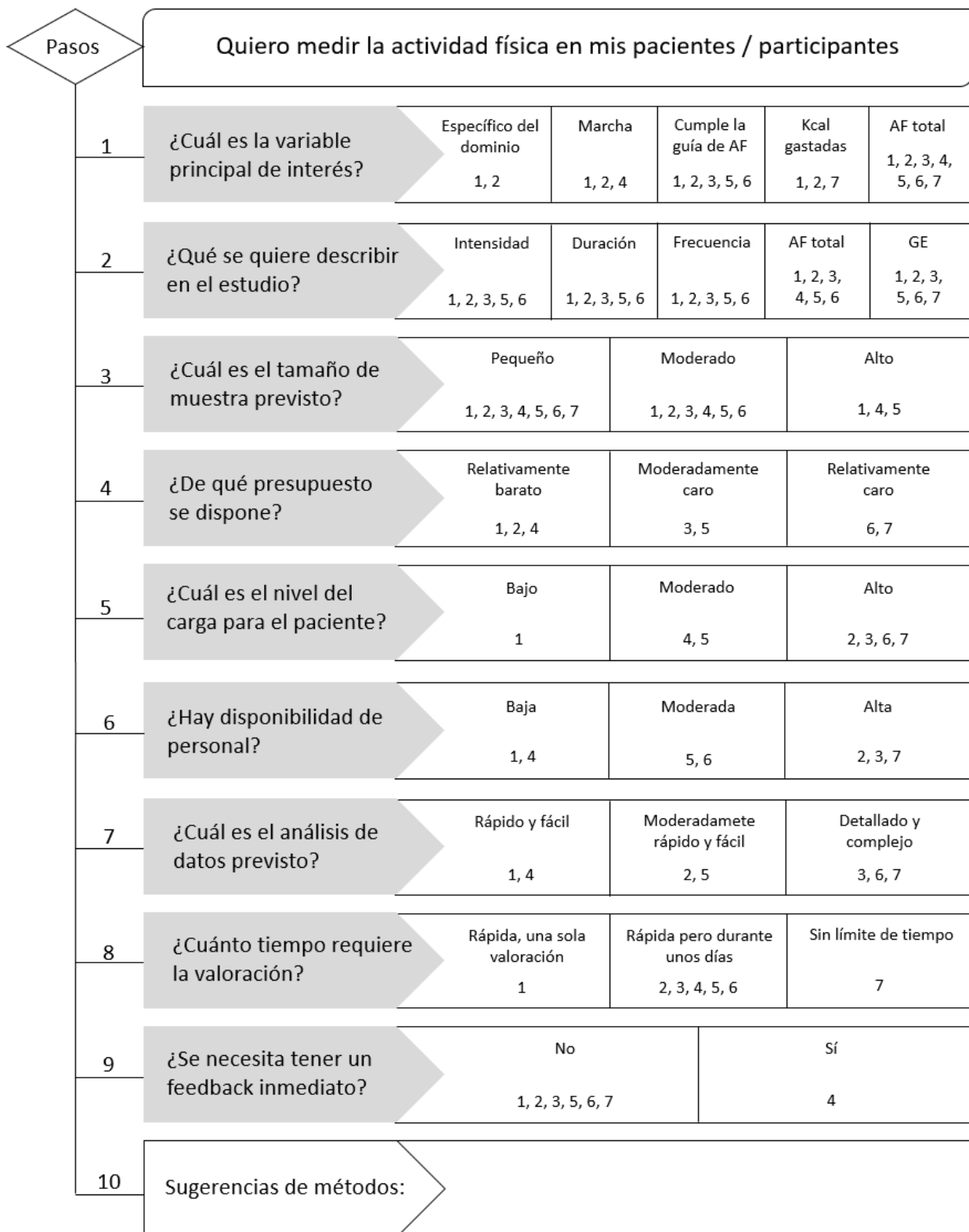


Figura 12. Algoritmo de decisión. Adaptación de Pettee Gabriel KK et al. 2012 (10)



1: cuestionario de actividad física; 2: diario de actividad física, 3: monitor de frecuencia cardíaca; 4: podómetro; 5: acelerómetro; 6: sensores de unidades múltiples; 7: agua doblemente marcada

Figura 13. Algoritmo de decisión. Adaptación de The American Heart Association. 2013 (58)

1.3.4. ¿Cómo escoger un cuestionario de actividad física?

Al igual que ocurre con los instrumentos de medición de la AF, la gran cantidad de cuestionarios disponibles dificulta la elección del más adecuado para un propósito específico (9,194,195). Esta elección se ve obstaculizada también por informes incompletos y una evaluación insatisfactoria del contenido de los cuestionarios (9,194,195). Por tanto, es primordial disponer de un sistema estandarizado de decisión que asegure que el cuestionario es apropiado para el objetivo de estudio y para las habilidades e intereses de la población diana (193).

Concretamente, en julio de 2010, se celebró una conferencia para explorar los desafíos y oportunidades de los métodos de autoinforme con el objetivo de crear un repertorio de cuestionarios de AF y alentar a los investigadores a desarrollar habilidades evaluativas para medir este comportamiento (196). A raíz de esta iniciativa, se han publicado diferentes bases de datos (197,198) y estudios de revisión que ofrecen directrices para evaluar la calidad de los cuestionarios de AF existentes, facilitar el proceso de selección y/o guiar la construcción de nuevos instrumentos autoinformados. Estas propuestas incluyen el checklist de Terwee et al., en el que se ofrece una estructura para evaluar la calidad de los atributos cualitativos y de las propiedades de medición (192), la guía práctica de Sternfeld et al. para definir los criterios importantes de las medidas indirectas de AF y CS (193) o el checklist de Hagströmer et al. para evaluar la calidad metodológica de los estudios sobre propiedades de medición (199).

Al igual que las matrices de decisión propuestas en el capítulo anterior, las guías que orientan en la elección de un cuestionario de AF consideran una gran variedad de factores intrínsecos y del contexto de aplicación. Estas pautas no pretenden ser prescriptivas, es decir, no asumen la potestad de decisión del investigador, sino que están destinadas a respaldar y ordenar el

proceso de elección de la forma más flexible y práctica posible. Cada uno de los atributos cualitativos y las propiedades de medición ponen de manifiesto los criterios específicos que debe cumplir un cuestionario para un propósito dado, pero ninguno de ellos tiene la capacidad de revelar la elección ideal de forma aislada. Sin embargo, una reflexión conjunta ayuda a reducir opciones potenciales y orienta hacia las más adecuadas desde un punto de vista crítico (192,193,200).

1.3.4.1. Atributos cualitativos

En general, los cuestionarios difieren en sus atributos cualitativos, como el constructo que miden, la población para la cual fueron diseñados, el período de recuerdo al que se remontan las preguntas, la claridad de las instrucciones introductorias, el número de preguntas, el número y tipo de opciones de respuesta, el algoritmo de puntuación, el objetivo de valoración, la posibilidad de interpretar clínicamente los resultados y las limitaciones logísticas asociadas a su aplicación, como el coste de tiempo, el modo de administración o la disponibilidad de una versión gratuita (192,193).

1.3.4.1.1. Constructo

El constructo se refiere a la descripción bien definida y precisa del concepto que un cuestionario pretende medir (201). Un constructo claramente informado es fundamental para diseñar un cuestionario adecuado a un objetivo, para evaluar la validez y, posteriormente, para facilitar la elección del más apropiado para un propósito específico (199,202).

Los cuestionarios de AF pueden estar diseñados para medir constructos diferentes, como el GE (203), el nivel de AF habitual (204–207), el nivel de AFMV (208,209), etc. Así mismo, pueden reportar la AF que ocurre en todos

los dominios (204,205,210) o solo en algunos de ellos, como el tiempo libre (206,211) o el transporte (212).

Al cuantificar la AF como parte del constructo que se pretende medir y dentro del dominio de interés, los datos recopilados deben incluir información acerca del tipo de actividad, la duración, la frecuencia y la intensidad (Figura 14). Sin embargo, la forma en la que se recoge la información de estas dimensiones puede variar en los diferentes cuestionarios.

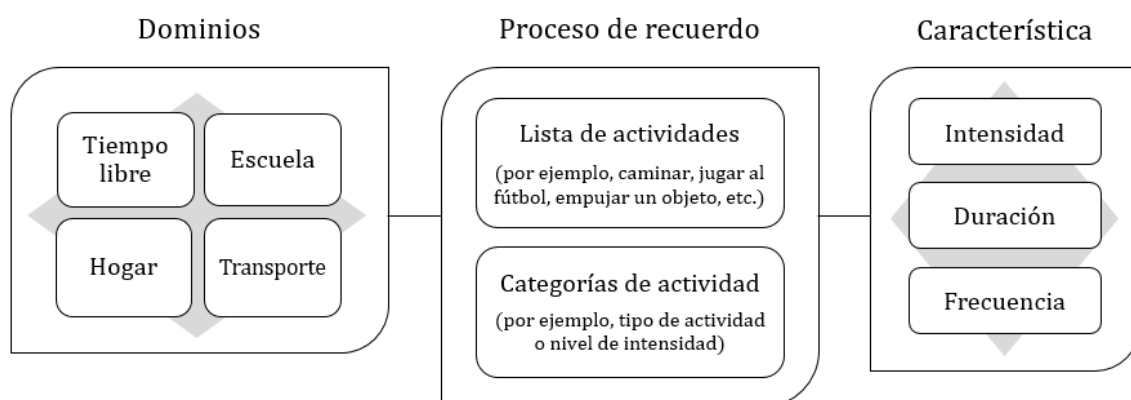


Figura 14. Proceso de recogida de información en los cuestionarios de actividad física. Adaptación de Pettee Gabriel KK et al. 2012 (20)

Los tipos de actividades que recogen los cuestionarios pueden hacer referencia a hechos concretos (por ejemplo, jugar a un videojuego, practicar natación, caminar hacia la escuela, etc.) o pueden estar clasificados por categorías sin concretar la acción específica (20). En el primer caso, habitualmente se ofrecen listas cerradas con diferentes actividades de interés para la población diana. Si bien este enfoque puede facilitar el proceso de recuperación de recuerdos, una limitación es que el cuestionario puede volverse relativamente largo y es posible que omita información importante para algunos de los encuestados (193). Por otro lado, los cuestionarios que piden información sobre categorías de actividad, ya sea por tipo o por

intensidad, pueden ser más completos al permitir la inclusión de una gama abierta de actividades, pero pueden ser más complejos debido a la demanda cognitiva para sumar el nivel de participación en todas las actividades relevantes de cada categoría (187).

Con respecto a la asignación de un nivel de intensidad, se han descrito discrepancias relacionadas con el enfoque en términos absolutos o relativos. Aunque ambas opciones están sujetas a errores de medición, se ha reconocido que la magnitud del error se puede minimizar planteando la actividad en términos de intensidad absoluta y asignando a cada actividad un valor MET específico obtenido de listas de referencia (capítulo 1.2.4.1) (102,114,117). Además, se ha descrito que los cuestionarios que se expresan en términos relativos, pasan por alto que la percepción de la intensidad depende en gran medida de la edad, el sexo y la aptitud física (213).

La duración y la frecuencia son otro desafío que medir ya que las personas se involucran constantemente en diferentes actividades de duraciones muy variables. En los cuestionarios de AF, las opciones de respuesta generalmente se dan como opciones de intervalo con minutos por día o por semana. Esto constituye una de las principales limitaciones, especialmente en niños, puesto que están sujetos al sesgo de memoria a la hora de registrar breves períodos de actividad, como subir las escaleras, coger objetos, desplazarse cortas distancias, etc. (56,100). Por este motivo, los cuestionarios de AF para niños deben diseñarse de tal manera que el impacto de las habilidades cognitivas y de memoria se reduzcan a un mínimo aceptable (214).

1.3.4.1.2. Población diana

La población diana a la que se dirige un cuestionario determina el contenido y su aplicabilidad. Por ejemplo, un cuestionario desarrollado para niños incluirá preguntas sobre juegos al aire libre, transporte activo a la escuela o

contenido lectivo relacionado con la Educación Física. Sin embargo, todos estos aspectos son irrelevantes en cuestionarios dirigidos a población adulta (177).

En general, la idoneidad de un cuestionario depende del contenido en cuanto a los dominios o tipos de actividades que se recogen y de los requerimientos cognitivos de los participantes. Además, se ha descrito que la edad, el género, la etnia, el nivel educativo, el idioma y el contexto cultural son factores personales que pueden condicionar la aplicación de un cuestionario a una población diferente a la original (172,177,193).

1.3.4.1.3. Período de recuerdo

El período de recuerdo de un cuestionario se refiere al rango de tiempo al que se remontan las preguntas. Como ya se comentó en el capítulo 1.3.2, las demandas cognitivas subyacentes al almacenamiento y recuperación de recuerdos son complejas, especialmente en niños (188,189). En consecuencia, las respuestas de los cuestionarios dependen del sesgo de memoria a la hora de definir con exactitud los patrones de AF (111,172,178). Además, la naturaleza esporádica de la AF de los niños hace que estas actividades sean aún más difíciles de recordar, cuantificar y categorizar en términos de frecuencia, duración e intensidad (15).

En algunos casos, los cuestionarios se remontan a la AF practicada durante un largo período de tiempo, como los últimos meses (207,215), el último año (206,216) o incluso toda la vida (217). La ventaja de este enfoque es que recupera la AF durante un período lo suficientemente amplio como para reducir al mínimo los eventos no representativos de lo considerado como habitual. Sin embargo, impone una gran carga cognitiva que obliga a promediar niveles variables de participación y conlleva un importante error de medición asociado al sesgo de memoria (218).

Un método alternativo para reducir la carga cognitiva y el error de medición es preguntar por un período de tiempo más cercano, como la última semana (208,219) o una semana habitual (205,220). Los desarrolladores del *International Physical Activity Questionnaire* (IPAQ) describieron que la interpretación de una semana habitual suele resultar compleja porque los participantes no son capaces de identificar lo que es habitual (203). Sin embargo, Terwee et al. defienden que no hay consenso sobre qué formato es más apropiado, sino que depende del constructo que se va a medir y del objetivo del estudio (192).

Por último, se han desarrollado cuestionarios con períodos de recuerdo muy recientes, como el último día (220,221). Si bien el error de medición y el sesgo de memoria pierden peso en este enfoque, se cuestiona la representatividad del comportamiento habitual en un período de recuerdo tan limitado (222).

1.3.4.1.4. Formato

El formato se refiere a la presentación y organización del cuestionario en lo que respecta al número de ítems, el número y tipo de opciones de respuesta y el algoritmo de puntuación (192).

Todos los cuestionarios deben ir acompañados de unas instrucciones introductorias que faciliten la comprensión de los objetivos y cautiven el interés del encuestado para responder a las preguntas con la máxima atención. Así mismo, la organización de los ítems ha de presentar una estructura clara y una secuencia lógica que guíen al participante a lo largo del cuestionario (200).

Con respecto al formato de los ítems, preferiblemente se recomiendan preguntas cerradas que expliquen claramente la información que se pretende recoger y formuladas con un lenguaje adaptado a las características de la

población para la cual fue diseñado (172). El número de preguntas varía desde cuestionarios de un solo ítem (204,220) a cuestionarios entre 20 a 60 preguntas (219,223).

En cuanto a las opciones de respuesta, algunos cuestionarios están diseñados con opciones cerradas que califican el nivel de AF en una escala Likert (208), mientras que otros proponen opciones semiabiertas que pretenden recoger información más precisa (224).

Dado que la mayoría de los cuestionarios de AF constan de varias preguntas, los ítems individuales generalmente se resumen en una o más variables que se utilizan en análisis posteriores. Dependiendo de los objetivos del estudio y del detalle de los ítems individuales, la variable resumen puede ser dicotómica (por ejemplo, activo o inactivo físicamente), ordinal categórica (por ejemplo, nivel de actividad bajo, medio o alto) o continua (por ejemplo, MET-horas/semana, tiempo dedicado a AFMV/semana o tiempo dedicado a una actividad específica) (193).

1.3.4.1.5. Objetivo

Las aplicaciones potenciales de los instrumentos de medición en el campo de la salud se pueden dividir en tres categorías principales: índices discriminativos, predictivos y evaluativos.

Los índices discriminativos se utilizan para clasificar a los individuos en una dimensión subyacente (225). En el ámbito de la AF, los cuestionarios discriminativos son aquellos diseñados para clasificar a las personas en grupos suficientemente activos o inactivos o evaluar la prevalencia y las diferencias en los patrones de AF (192).

Los índices predictivos sitúan a los individuos en un conjunto de categorías de medición predefinidas. Este tipo de índices se usan generalmente como un

instrumento pronóstico de detección o diagnóstico para identificar qué individuos específicos tienen o desarrollarán una condición o resultado (225). En el campo de la AF, los cuestionarios predictivos deberían ser capaces de pronosticar la relación entre la AF y otros indicadores de salud (192).

El objetivo de los índices evaluativos es medir la magnitud de un cambio longitudinal en la dimensión de interés en un individuo o un grupo. Los cuestionarios de AF con fines evaluativos sirven para monitorear los patrones de AF a lo largo del tiempo o evaluar el efecto de una intervención terapéutica (192,225).

En la construcción de un cuestionario de AF es importante definir a priori el objetivo para ajustar los requerimientos de validación. Aun así, el proceso de desarrollo de la mayoría de los instrumentos es idéntico independientemente de si se diseña con fines discriminativos, evaluativos o predictivos. En otras palabras, un cuestionario validado con un propósito específico no indica necesariamente que no sea útil para los otros, pero se ha de demostrar mediante el estudio de las propiedades de medición pertinentes (225). Por ejemplo, en un cuestionario con fines discriminativos el dominio fiabilidad es muy importante mientras que en un cuestionario diseñado para evaluar el efecto de una intervención, la capacidad de respuesta es más relevante (192,225).

1.3.4.1.6. Interpretabilidad

La interpretabilidad se define como el grado en el que se puede asignar un significado cualitativo a las puntuaciones cuantitativas de un instrumento, específicamente, connotaciones clínicas o comúnmente entendidas (226).

Aunque algunas puntuaciones de AF tienen un significado intuitivo debido a que se expresan en parámetros conocidos como MET-horas/semana, tiempo dedicado a AFMV o tiempo dedicado a una actividad específica, las puntuaciones adimensionales son más difíciles de interpretar. En tal caso, es de gran utilidad disponer de valores de normalidad de la población general según diferentes grupos de edad y sexo para tomarlos como valores de referencia y facilitar su comparación, por ejemplo, con los resultados sobre los patrones de AF en poblaciones patológicas (192).

Por otro lado, es útil saber cuál es el cambio mínimo en la puntuación del cuestionario que se corresponde con un cambio significativo en la AF, es decir, el cambio mínimo importante. Este valor es útil para la interpretación del error de medición y para los cálculos del tamaño de la muestra (192).

1.3.4.1.7. Limitaciones logísticas

La evaluación de los niveles habituales de AF está sujeta a restricciones presupuestarias y relacionadas con la carga del encuestado y del equipo investigador (192,193).

Estas limitaciones afectan a la viabilidad de los diferentes modos de administración, entre los que se encuentra el auto informe, el informe por parte de un cuidador o la administración por un entrevistador capacitado. Así mismo, se debe valorar la vía de difusión de los cuestionarios, que puede ser en persona, por teléfono, por correo postal o a través de una plataforma online (168). Todas estas decisiones condicionan la cantidad de tiempo y esfuerzo que pueden dedicar tanto los participantes como los investigadores y, en ocasiones, limitan la longitud y complejidad del instrumento seleccionado a favor de que sea práctico y presente buena aceptabilidad (193).

Aunque estos aspectos son fundamentales para garantizar el éxito de cualquier proyecto, con demasiada frecuencia son el factor determinante en la elección de un instrumento de AF (193).

1.3.4.2. Propiedades de medición

Los cuestionarios también se diferencian en sus propiedades de medición, es decir, su validez, su fiabilidad y su capacidad de respuesta. Además, la calidad de los estudios realizados para probar las propiedades de medición puede interferir en los resultados obtenidos. Por ejemplo, la muestra del estudio debe ser representativa de la población a la que va dirigido el cuestionario, el tamaño de muestra seleccionado debe ser suficiente, el análisis estadístico realizado debe ser apropiado, etc. (177,202). Por tanto, la elección de un cuestionario de AF es un equilibrio entre los atributos definidos en su diseño y las propiedades de medición demostradas con estudios de investigación (201).

Para comprender el impacto de las propiedades de medición en los cuestionarios de AF, es importante entender el marco conceptual que explica la relación entre los ítems y el constructo que se va a medir (201). En 1997, Fayers et al. introdujeron la distinción entre los modelos reflexivos y los modelos formativos (227). En su forma más simple, su propuesta se representa en la Figura 15.

A la izquierda, se observa el esquema propio de un enfoque reflexivo. La representación sugiere que el constructo se manifiesta en los ítems, es decir, el constructo es el reflejo directo de los ítems. Por tanto, los ítems funcionan como indicadores del efecto. A la derecha, el modelo formativo establece una relación en la que el constructo es el resultado de los ítems. En otras palabras, los ítems forman o causan el constructo y funcionan como indicadores causales (201).

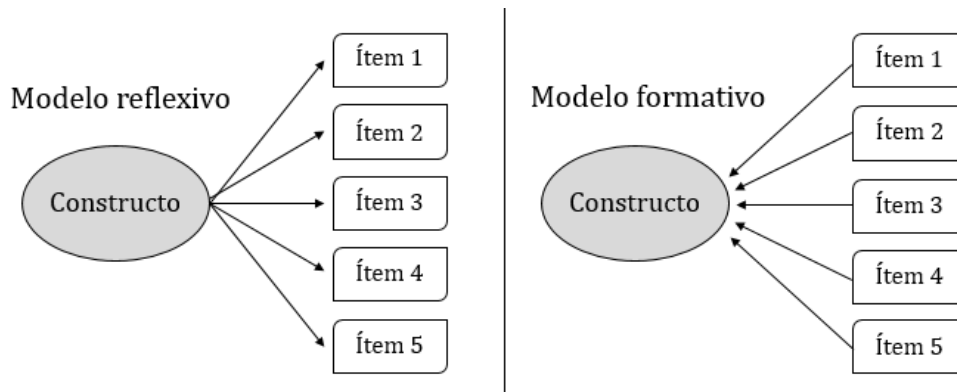


Figura 15. Representación de un modelo reflexivo y un modelo formativo. Adaptación de Vet et al. 2011 (201)

La forma más sencilla para decidir si la relación entre los ítems y el constructo se basa en un modelo reflexivo o formativo es realizar una “prueba de pensamiento”. Si se espera que un cambio en el constructo influya en todos los ítems, se trata de un enfoque reflexivo. Por el contrario, si el cambio en el constructo no afecta necesariamente a todos los ítems, el modelo probablemente sea formativo. En este tipo de enfoque no es posible predecir la dirección de todos los ítems cuando el constructo varía, sino que hay que analizar cada ítem de forma individual (201).

Los cuestionarios de AF claramente están basados en un modelo formativo. Un cambio en el nivel de AF no implica necesariamente un cambio en todos los ítems del cuestionario. Por ejemplo, es posible que un individuo logre aumentar sus niveles de AF cambiando los desplazamientos en coche por una rutina de transporte activo, pero sin modificar su conducta en el tiempo libre (192).

La construcción de un cuestionario bajo un modelo reflexivo o formativo determina también la teoría de medición que sustenta su desarrollo y evaluación. La teoría de medición pretende explicar cómo las puntuaciones generadas por los ítems representan el constructo que se va a medir (228). Las dos teorías más conocidas son la Teoría Clásica de Medición y la Teoría de

Respuesta al Ítem, ambas aplicadas a modelos reflexivos. Para los cuestionarios basados en modelos formativos no se han planteado teorías de medición que expliquen las relaciones entre el constructo y el ítem (228). Esto no quiere decir que no exista un proceso subyacente a la construcción de un cuestionario de AF, sino que no hay una teoría claramente descrita que apunte la relación entre los ítems y el constructo (201).

Una vez entendido que los cuestionarios de AF están basados en un modelo formativo y su construcción no está explicada por ninguna teoría de medición, surge un nuevo reto con respecto a la discrepancia en la terminología y definiciones de las propiedades de medición.

A lo largo de los años, se ha observado una falta de consenso entre los términos y conceptos aplicados a estas características. Por ejemplo, en la literatura se utiliza indistintamente terminología diferente para la propiedad “fiabilidad”, como reproducibilidad (177,217,229), repetitividad (203,230,231), acuerdo (232–234) o estabilidad (235–237). Así mismo, se puede encontrar la capacidad de respuesta definida como *“la capacidad de un instrumento para detectar cambios en los resultados que son importantes para las personas con un problema de salud, sus seres queridos o sus proveedores”* (226) o como *“la capacidad de un instrumento para detectar cambios a lo largo del tiempo en el constructo que se va a medir”* (238). Mientras que los diferentes usos de la terminología pueden generar confusión sobre qué propiedad de medición se evalúa, las diferencias en las definiciones pueden llevar a la confusión acerca de qué concepto representa la propiedad de medición y cómo debe evaluarse. Por tanto, antes de estandarizar la selección de medidas de resultado, es necesario aclarar la taxonomía y definiciones sobre las propiedades de medición ya pueden conducir a nuevos sesgos sobre las conclusiones de los estudios (239).

Varias iniciativas, como El Comité Asesor Científico del Medical Outcomes Trust (SAC-MOS) (240), La Asociación Americana de Psicología (APA) (241)

o Terwee et al. (242) presentaron propuestas para identificar las propiedades de medición, pero estaban centradas en pruebas educativas y psicológicas (241) o no se basaban en el consenso de expertos (240,242).

En el 2010, la iniciativa COSMIN (*CO*n*SENSUS*-*BASED* *STANDARDS* *FOR* *THE* *SELECTION* *OF* *HEALTH* *MEASUREMENT* *INSTRUMENTS*) desarrolló una consulta mediante el método Delphi para estandarizar la terminología (Figura 16) y definiciones (Tabla 2) de las propiedades de medición. Tras cuatro rondas de debate, identificaron tres dominios: validez, fiabilidad y capacidad de respuesta, cada uno de ellos con una o más subcategorías, es decir, propiedades de medición o aspectos importantes de las propiedades de medición. El dominio “validez” contiene tres propiedades de medición: validez de contenido, validez de criterio y validez de constructo. El dominio “fiabilidad” contiene otras tres propiedades de medición: consistencia interna, fiabilidad y error de medición. Por último, el dominio “capacidad de respuesta” contiene solo una propiedad de medición, que también se denomina capacidad de respuesta. Algunas propiedades de medición abarcan además uno o más aspectos que se definen por separado. Por ejemplo, la validez de contenido incluye la validez aparente, mientras que la validez de constructo engloba la validez estructural, la prueba de hipótesis y la validez transcultural. Por otro lado, los expertos coincidieron en que la interpretabilidad es una característica relevante, aunque no la identificaron como una propiedad de medición, sino más bien como un atributo cualitativo (capítulo 1.3.4.1.6) (239).

La terminología y las definiciones de la iniciativa COSMIN difieren ligeramente de las propuestas comentadas anteriormente (239). Por ejemplo, la APA no incluye el dominio capacidad de respuesta. En psicología, las medidas suelen usarse con fines discriminativos y, por tanto, la capacidad de respuesta no es relevante (241). Por otro lado, Terwee et al. no consideran la consistencia interna como una subcategoría del dominio fiabilidad (242). Así

mismo, en la propuesta de SAC-MOS no se incluye el error de medición en la descripción general de las propiedades de medición (240).

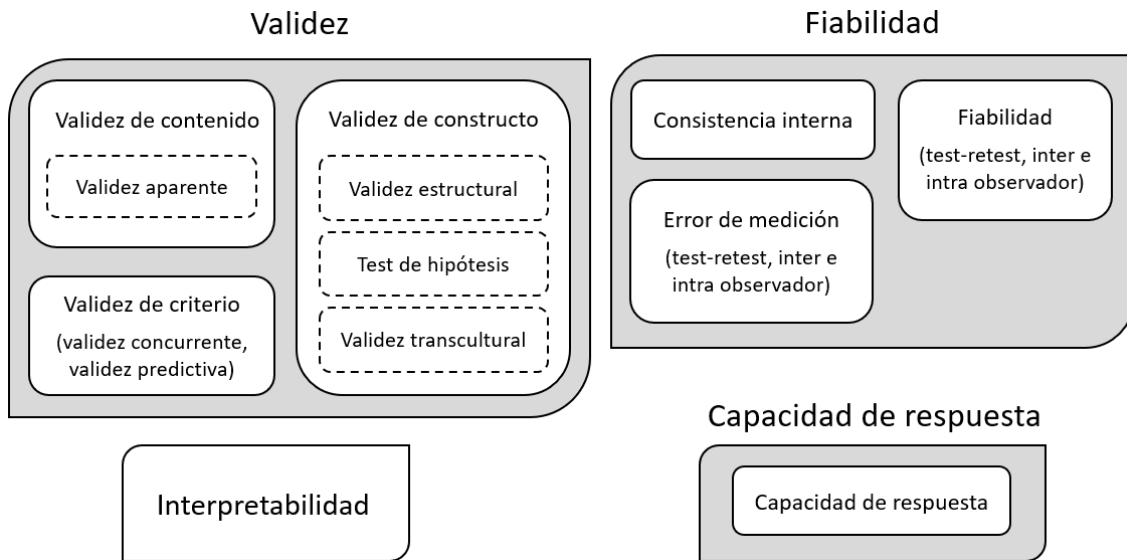


Figura 16. Taxonomía COSMIN de las propiedades de medición. Adaptación de Mokkink et al. 2010 (239)

Tabla 2. Definiciones COSMIN de las propiedades de medición. Adaptación de Mokkink et al. 2010 (239)

Dominio	Propiedad	Aspecto	Definición
Validez			El grado en que un cuestionario mide el constructo que pretende medir.
		Validez de contenido	El grado en que el contenido de un cuestionario es un reflejo adecuado del constructo a medir.
		Validez aparente	El grado en que los ítems de un cuestionario realmente son un reflejo adecuado del constructo a medir.
		Validez de criterio	El grado en que las puntuaciones de un cuestionario son un reflejo adecuado de un <i>gold standard</i> .
		Validez de constructo	El grado en que las puntuaciones de un cuestionario son coherentes con las hipótesis basadas en el supuesto de que el cuestionario mide válidamente el constructo a medir.
		Validez estructural	El grado en que las puntuaciones de un cuestionario son un reflejo adecuado de la dimensionalidad del constructo a medir.
		Test de hipótesis	Ídem a la validez de constructo.
	Validez transcultural	El grado en que la adaptación de los ítems en un cuestionario traducido o adaptado culturalmente es un reflejo adecuado de la representación de los ítems de la versión original.	
Fiabilidad			El grado en que la medición está libre de error de medición.
		Consistencia interna	El grado de interrelación entre los ítems.
		Fiabilidad	La proporción de la varianza total en las mediciones que se debe a diferencias "verdaderas" entre los pacientes.
		Error de medición	El error sistemático y aleatorio de la puntuación de un individuo que no se atribuye a cambios verdaderos en el constructo a medir.
Capacidad de respuesta			La capacidad de un cuestionario para detectar cambios a lo largo del tiempo en el constructo a medir.
		Capacidad de respuesta	Ídem a la capacidad de respuesta.

1.3.4.2.1. Validez de contenido

La validez de contenido se define como el grado en el que el contenido de un instrumento es un reflejo adecuado del constructo a medir (239). Se considera la propiedad de medición más importante y a la vez la más difícil de determinar (243,244).

Como punto de partida, es importante disponer de una descripción suficientemente clara del constructo para juzgar si los ítems son relevantes y si el constructo queda cubierto de manera integral. Una de las mayores limitaciones para la validez de contenido es una coincidencia difusa entre el cuestionario y el constructo que se pretende medir (245). En este contexto, si fuera posible, se debería aportar información adicional sobre el marco o modelo teórico en el que se fundamenta el cuestionario. Además, es fundamental describir para quién se desarrolló el cuestionario, cuál es el contexto de aplicación y cuál es el objetivo de evaluación previsto. Todos estos aspectos condicionan la especificidad y amplitud de los ítems para determinar la aplicabilidad del cuestionario en circunstancias diferentes a la prevista originalmente (243,244).

La forma más extendida de evaluar la validez de contenido es a través de entrevistas con profesionales y muestras representativas de la población diana para indagar sobre la relevancia, exhaustividad y comprensibilidad de cada uno de los ítems. En primer lugar, todos los ítems del cuestionario deben ser relevantes para el constructo de interés dentro de la población, el contexto y el objetivo de uso específico. En segundo lugar, el cuestionario debe ser lo suficientemente exhaustivo como para asegurar que se recogen todos los aspectos claves del constructo. En tercer lugar, los ítems deben ser fácilmente comprendidos por la población diana en el sentido en el que han sido formulados (243,244). En los cuestionarios de AF, se considera prioritario evaluar también la cantidad de detalles con respecto a la frecuencia, la duración, la intensidad y el tipo de actividades incluidas para determinar si

son apropiadas para el constructo de interés, la población, el contexto, el objetivo de evaluación y el período de recuerdo (192).

La validez aparente se incluye como una característica de la validez de contenido. En general, pretende arrojar luz sobre la idoneidad de las características evidentes de un cuestionario. A diferencia de la validez de contenido, que evalúa el detalle de los ítems, la validez aparente evoca una reflexión sobre aspectos globales, por ejemplo, si los ítems invitan a una respuesta precisa, si la combinación de ítems y puntuaciones es lógica, si la información se pide de manera apropiada o si la extensión y el formato del cuestionario son adecuados para la población diana (192). Sin embargo, a menudo recibe poca atención porque se basa en un juicio subjetivo sobre el contenido del cuestionario que no puede cuantificarse estadísticamente ni compararse con un estándar predefinido. Por tanto, su valor como índice de la validez de contenido es indeterminado (201).

Algunos autores consideran también los efectos techo y suelo como parte del estudio de validez de contenido de un cuestionario (192,201). Los efectos techo y suelo se definen como la distribución de las puntuaciones de manera que más del 15% de los respondedores muestran la puntuación más alta o baja posible, respectivamente (246). En otras palabras, si muchos individuos obtienen la mínima o la máxima puntuación, no pueden distinguirse entre sí, por lo que la fiabilidad es cuestionable. Así mismo, la calidad de la capacidad de respuesta es pobre ya que el cuestionario no sería capaz de detectar cambios en la AF en las personas que tienen la puntuación más alta o más baja (192). Estos efectos se deben considerar desde las fases de desarrollo del cuestionario ya que evidencian un problema en la selección de los ítems y obligan a revisar el contenido (201,247).

1.3.4.2.2. Validez de criterio

El nivel más alto de evidencia de validez se obtendría comparando el cuestionario de AF con un *gold standard*, es decir, con un instrumento que mide el mismo constructo y tiene perfecta fiabilidad y validez (192,193).

Tal y como se ha comentado anteriormente (ver capítulo 1.3.3) no existe un *gold standard* real en la medición de la AF (20,111,177,190–192). Además, el panel de COSMIN llegó al consenso de que tampoco existe un *gold standard* para los cuestionarios. La única excepción es cuando se compara un instrumento abreviado con la versión larga original. En ese caso, la versión larga original puede considerarse el instrumento de referencia (247).

Como conclusión, cuando se comparan las puntuaciones de un cuestionario con un instrumento ampliamente utilizado y conocido, se considera un estudio sobre validez de constructo (192,247).

1.3.4.2.3. Validez de constructo

La validez de constructo es aplicable en situaciones en las que no hay un *gold standard* para valorar si el cuestionario proporciona las puntuaciones esperadas, basadas en el conocimiento sobre el constructo (201). Aunque los resultados son menos concluyentes en comparación con la validez de criterio, si se fundamenta en modelos teóricos específicos y se plantean las hipótesis correctamente, es posible que ofrezca un nivel de evidencia suficiente como para afirmar que el cuestionario mide el constructo que pretende (201).

La validez de constructo presenta tres aspectos: el test de hipótesis, la validez estructural y la validez transcultural (239).

El principio básico de la validez de constructo consiste en probar hipótesis previamente formuladas sobre las relaciones entre las puntuaciones del

cuestionario en estudio y otros instrumentos que miden constructos similares u otros subgrupos conocidos (201,247). En los cuestionarios de AF, la validez de constructo se evalúa comparando las puntuaciones con otros instrumentos que miden constructos estrechamente relacionados. Cuanto más similares sean los constructos, más pruebas se proporcionarán para su validez (202). En las Tabla 3 y Tabla 4 se muestran las propuestas de Terwee et al. (192) y Hidding et al. (13) con respecto a las correlaciones esperadas para la validez de constructo de los cuestionarios de AF. En la Tabla 5 se presenta la adaptación de Hidding et al. para los cuestionarios que miden CS (248).

Tabla 3. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios de actividad física. Adaptación de Terwee et al. 2010. (192).

Constructo	Instrumento comparador	r
Gasto energético total	Agua doblemente marcada	$r \geq 0,7$
Actividad física total	Acelerómetro (recuentos totales)	$r \geq 0,5$
Actividad física vigorosa	Acelerómetro (tiempo de actividad vigorosa)	$r \geq 0,5$
Actividad física moderada	Acelerómetro (tiempo de actividad moderada)	$r \geq 0,5$
Marcha	Podómetro o acelerómetro (durante la marcha)	$r \geq 0,7$
Actividad física en el tiempo libre	Acelerómetro (durante el tiempo libre)	$r \geq 0,5$
Actividad física ocupacional	Observación directa	$r \geq 0,6$

r: coeficiente de correlación

Tabla 4. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios de actividad física. Niveles de evidencia 1, 2 y 3. Adaptación de Hidding 2018. (13)

Constructo	Nivel de evidencia 1	Nivel de evidencia 2	Nivel de evidencia 3
Actividad física: todos los dominios	Observación directa $\geq 0,7$ Recuentos totales de acelerómetro $\geq 0,6$ Agua doblemente marcada $\geq 0,6$	Recuentos de AFV, AFMV o AFM de acelerómetro $\geq 0,4$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,4$
Actividad física: algunos dominios o franjas	Observación directa $\geq 0,7$ Recuentos de acelerómetro en la franja $\geq 0,6$	Recuentos totales de acelerómetro $\geq 0,4$ Recuentos de AFMV de acelerómetro $\geq 0,5$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,4$
Actividad física: dominios aislados		Recuentos totales de acelerómetro $\geq 0,4$ Recuentos de AFMV de acelerómetro $\geq 0,5$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,4$ <i>Cycle computer</i> ^b $\geq 0,7$
Gasto energético	Agua doblemente marcada $\geq 0,7$	Recuentos totales de acelerómetro $\geq 0,5$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,4$
Actividad física vigorosa	Recuentos de AFV de acelerómetro $\geq 0,6$	Recuentos totales de acelerómetro $\geq 0,4$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,6$
Actividad física moderada-vigorosa	Recuentos de AFMV de acelerómetro $\geq 0,6$	Recuentos totales de acelerómetro $\geq 0,4$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,6$
Actividad física moderada	Recuentos de AFM de acelerómetro $\geq 0,6$	Recuentos totales de acelerómetro $\geq 0,4$ Recuentos de podómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$ $VO_{2m\acute{a}x} \geq 0,5$
Marcha	Recuentos de podómetro o de marcha de acelerómetro $\geq 0,7$	Recuentos totales de acelerómetro $\geq 0,4$	MI con mismos constructos $\geq 0,7$

AFV: actividad física vigorosa, AFMV: actividad física moderada-vigorosa, AFM: actividad física moderada, MI: método indirecto (cuestionario, diario o entrevista), $VO_{2m\acute{a}x}$: consumo máximo de oxígeno, ^b: Si se usa como comparación para el ciclismo

Tabla 5. Correlaciones esperadas e instrumentos recomendados para la validez de constructo de los cuestionarios que miden el comportamiento sedentario. Niveles de evidencia 1, 2 y 3. Adaptación de Hidding 2017. (248)

Constructo	Nivel de evidencia 1	Nivel de evidencia 2	Nivel de evidencia 3
Comportamiento sedentario: todos los dominios	ActivPAL $\geq 0,7$ Observación directa $\geq 0,7$	Acelerómetro 100 cpm $\geq 0,6$	MI con mismos constructos $\geq 0,7$ Acelerómetro $< 0 > 100$ cpm $\geq 0,4$
Sedestación (tiempo total)	ActivPAL $\geq 0,7$ Observación directa $\geq 0,7$	Acelerómetro 100 cpm $\geq 0,5$	MI con mismos constructos $\geq 0,7$ Acelerómetro $< 0 > 100$ cpm $\geq 0,4$
Tiempo con pantallas	Observación directa $\geq 0,7$	Diario $\geq 0,6$ Dispositivo de monitoreo de TV $\geq 0,6$	MI con mismos constructos $\geq 0,7$ Acelerómetro $\geq 0,4$
Comportamiento sedentario: algunos dominios o franjas	ActivPAL $\geq 0,7$ Observación directa $\geq 0,7$	Acelerómetro 100 cpm $\geq 0,6$ ^a Acelerómetro 100 cpm $\geq 0,5$ ^b	MI con mismos constructos $\geq 0,7$ Acelerómetro $< 0 > 100$ cpm $\geq 0,4$

cpm: recuentos por minuto, MI: métodos indirectos, ^a El período de recuerdo del cuestionario coincide con el período de medición del acelerómetro, ^b El período de recuerdo del cuestionario coincide con el período de medición del acelerómetro

En los cuestionarios de AF dirigidos a población general, no es frecuente encontrar comparaciones por subgrupos, excepto en los estudios de traducción y adaptación transcultural. La mayoría de las medidas de AF disponibles se han desarrollado en comunidades blancas de clase media, lo que presenta un desafío para los investigadores que quieren administrarlas a otros grupos étnicos o socioeconómicos (177,249). Si se producen modificaciones en el instrumento, como traducciones, adecuaciones culturales o cambios en el orden de las preguntas, se aconseja implementar un proceso de adaptación que explique cómo y por qué ocurrieron los cambios y que garantice que las modificaciones no influyan negativamente en la validez del cuestionario (249). Cuando se comparan las puntuaciones de un cuestionario entre dos subgrupos, es imprescindible informar sobre las características importantes para que quede claro en qué aspectos difieren, por ejemplo, edad, género, características de la enfermedad, idioma, etc. (250).

Por último, la validez estructural se considera un análisis propio de la estructura interna. Estos análisis solo son relevantes para los cuestionarios que se basan en un modelo reflexivo ya que se asume que todos los ítems son manifestaciones del constructo y se espera que estén correlacionados. Por el contrario, los cuestionarios diseñados bajo modelos formativos no contemplan una correlación de los ítems, por lo que los análisis de la estructura son irrelevantes (247). Tal y como se ha comentado al inicio del capítulo 1.3.4.2, los cuestionarios de AF están basados en modelos formativos en los que no se espera una homogeneidad en los ítems (9,192).

1.3.4.2.4. Consistencia interna

La consistencia interna representa el grado de interrelación entre los ítems (239). Al igual que la validez estructural, se considera una propiedad de

medición que caracteriza la estructura interna, por lo que carece de importancia en los cuestionarios de AF (9,192).

1.3.4.2.5. Fiabilidad

La fiabilidad es el grado en que la medida está libre de error de medición, es decir, el grado en que las puntuaciones de los respondedores son las mismas en mediciones repetidas siempre y cuando se hayan mantenido estables las condiciones de medición (239). La fiabilidad es un requerimiento mínimo de todas las medidas utilizadas tanto en la práctica clínica como en investigación, pero a veces no se reconoce su importancia hasta que se realizan las mediciones test-retest (201).

La fiabilidad debe ser evaluada en dos administraciones consecutivas (fiabilidad test-retest), ya sean del mismo respondedor o evaluador (intraobservador) o de entrevistadores o evaluadores diferentes (fiabilidad interobservador). Las administraciones deben ser independientes entre sí y realizarse en las mismas condiciones, es decir, se debe asegurar que la primera administración no ha influido en la segunda en el sentido en que el sujeto no debe ser consciente de las puntuaciones que ha obtenido (247). En otras palabras, el estudio de fiabilidad debe proporcionar estimaciones razonablemente similares de AF con mediciones repetidas del mismo individuo, asumiendo un efecto mínimo de aprendizaje y un cambio despreciable en el comportamiento a lo largo del tiempo (193).

El intervalo de tiempo entre el test y el retest es todavía objeto de discusión entre los expertos y no hay un claro consenso sobre los tiempos recomendados. El período entre la primera y la segunda medición debe ser lo suficientemente largo para asegurar que el individuo olvide las respuestas del primer test, pero lo suficientemente corto para garantizar que los patrones de AF no hayan cambiado (177). En este sentido, Chinapaw et al. proponen los

siguientes criterios: entre 1 día y 3 meses para los cuestionarios con período de recuerdo de “una semana habitual”, entre 1 día y 2 semanas para los cuestionarios con período de recuerdo de “la última semana” y entre 1 día y 1 semana para los cuestionarios con período de recuerdo del “último día”. En los dos últimos casos, se tiene en consideración además que el test y el retest hagan referencia a los mismos días, sin tener en cuenta el tiempo transcurrido entre las dos administraciones (9).

1.3.4.2.6. Error de medición

El error de medición es el error sistemático y aleatorio de la puntuación de un sujeto que no se atribuye a cambios verdaderos en el constructo que se va a medir (239).

El error de medición pone de manifiesto que cualquier observación se compone de una puntuación verdadera y un error asociado con la observación (247). Representa la idea de que las medidas realizadas en contextos clínicos o de investigación no muestran la puntuación real, sino que están sometidas a variaciones derivadas de diferentes fuentes, como los instrumentos, los evaluadores, los sujetos sometidos a la medida o las circunstancias bajo las que se realizan las valoraciones (201,222). Este principio se cumple para todas las mediciones, ya sean directas o indirectas, sobre un modelo reflexivo o formativo, etc. (201) y se expresa mediante estadísticos que evalúan cómo de cerca están las puntuaciones en administraciones repetidas (251).

La estimación del error de medición está íntimamente relacionada con la evaluación de la fiabilidad test-retest de un cuestionario. El protocolo de valoración para estudio del error de medición debe cumplir los mismos requisitos que para el estudio de fiabilidad, es decir, debe ser independiente,

debe realizarse en las mismas circunstancias y debe respetar un intervalo de tiempo adecuado entre el test y el retest (9,200,247).

1.3.4.2.7. Capacidad de respuesta

La capacidad de respuesta es la capacidad de un instrumento para detectar cambios a lo largo del tiempo en el constructo que se va a medir. Aunque la capacidad de respuesta se considera una propiedad de medición independiente del dominio validez, en el Delphi desarrollado por la iniciativa COSMIN se llegó al consenso de que la única diferencia entre la validez (de constructo o de criterio) y la capacidad de respuesta es que la primera se refiere a una única puntuación y la segunda afecta a una puntuación de cambio (84% de acuerdo en la consulta Delphi) (239).

Al evaluar la capacidad de respuesta, el objetivo es demostrar si el instrumento mide correctamente los cambios en el constructo, pero también si es capaz de determinar la cantidad correcta de cambio, es decir, si valora la magnitud esperada del efecto del tratamiento sin subestimar ni sobrestimar el cambio real. Este último aspecto constituye uno de los mayores retos en el estudio de la capacidad de respuesta, aunque menudo se pasa por alto (247).

La capacidad de respuesta debe ser evaluada por dos administraciones del cuestionario. Entre las administraciones, al menos algunos de los sujetos deberían haber cambiado sus niveles de AF en un grado relevante. De manera análoga a la validez de constructo, la capacidad de respuesta se puede evaluar comparando los cambios en el cuestionario de AF antes y después de una intervención con los cambios en otros instrumentos que miden constructos estrechamente relacionados. Otro posible enfoque es examinar cómo el cuestionario puede distinguir entre personas que han cambiado sus niveles de AF y personas que se han mantenido estables, basándose en algún criterio externo, por ejemplo comparando un grupo de entrenamiento con un grupo

control (192,247). Al igual que en los estudios de validez, cuando se comparan las puntuaciones de un cuestionario de AF con un instrumento ampliamente utilizado y conocido, se considera un estudio sobre capacidad de respuesta bajo el enfoque de constructo y se descartan los estudios bajo el enfoque de criterio (192,247).

1.3.4.3. Cuestionarios dirigidos a edad preescolar y escolar

Cada uno de los atributos y las propiedades de medición explicadas hasta el momento ponen de manifiesto los aspectos relevantes en la elección de un cuestionario de AF. Si bien es cierto que las características cualitativas pueden condicionar la aplicación de un cuestionario, los expertos coinciden en que el estudio de las propiedades de medición es un requisito indispensable de todos los cuestionarios (192,193,200).

En este contexto, las propiedades de medición se han estudiado de forma deficiente para muchos de los cuestionarios destinados a la población general. Los métodos utilizados varían tanto en contenido como en calidad y la mayoría de los estudios presentan importantes limitaciones metodológicas, como un tamaño de muestra insuficiente, deficiencias en el diseño, análisis estadísticos inadecuados, falta de coherencia con respecto al enfoque utilizado para medir la AF o presentaciones incompletas de los informes que describen el desarrollo y la validez de contenido (9,192,194,195).

En relación a los cuestionarios que valoran los niveles de AF en edad preescolar y escolar, se han publicado varias revisiones sistemáticas que analizan los estudios sobre propiedades de medición. En general, los cuestionarios dirigidos a esta franja de edad varían según el constructo que se pretende medir, el rango de edad al que se dirigen, el período de recuerdo al que se remontan las preguntas, el número de preguntas, el número y tipo

de opciones de respuesta, el coste de tiempo y el modo de administración (192,193).

En el 2000, se realizó la primera revisión sistemática sobre propiedades de medición de cuestionarios de AF dirigidos a niños y adolescentes. Aunque ya se habían publicado otras revisiones (14,252), hasta el momento no se había ofrecido una síntesis completa de todos los cuestionarios disponibles. Los resultados de esta revisión demostraron correlaciones de gran variabilidad para la validez de constructo ($r = 0,03 - 0,88$) y resultados muy dispares con respecto a la fiabilidad test-retest (ICC o Kappa = $0,2 - 0,99$) (7).

En 2008, se publicó otra revisión sistemática similar a la anterior (8), en la que se proponía el *Physical Activity Questionnaire - Older Children* (PAQ-C) como la mejor opción disponible en cuanto a fiabilidad (ICC = $0,75 - 0,82$) y validez de constructo ($r = 0,41 - 0,63$) (208,253).

En el 2010, miembros de la iniciativa COSMIN llevaron a cabo otra revisión sistemática sobre las propiedades de medición de los cuestionarios de AF dirigidos a menores de 18 años. Incluyeron 6 cuestionarios para evaluar la AF en niños en edad preescolar y 25 cuestionarios para niños entre 6 y 12 años (9). Con respecto a la validez de constructo, en edad preescolar, la correlación más alta se encontró para el *Children's Physical Activity Questionnaire* (CPAQ) ($r = 0,42$) (219) y para el *Netherlands Physical Activity Questionnaire for Young Children* (NPAQ) ($r = 0,33$ y $0,36$ para la AF total y la AF vigorosa, respectivamente) (215). En niños entre 6 y 12 años, las correlaciones más altas se reportaron para el *Physical Activity Questionnaire for Parents and Teachers* ($r = 0,53$) (210) y para el *ACTIVITY* ($r = 0,40$) (254). Con respecto a la fiabilidad, en preescolares, el cuestionario *Children's Leisure Activities Study Survey* (CLASS) fue el que mostró mejores resultados (ICC = $0,49-0,87$) (205), mientras que en los niños, los cuestionarios más fiables fueron el *Girls Health Enrichment Multisite Study Activity Questionnaire* (GAQ) (ICC = $0,82$) (255) y el PAQ-C (ICC = $0,75 - 0,82$) (253).

En 2012, Foley et al. desarrollaron una revisión sistemática para evaluar la calidad de los cuestionarios de AF y CS que provocan un recuerdo retrospectivo estructurado y ordenado cronológicamente durante un período de tiempo delimitado. Ninguno de los cuestionarios analizados estaba dirigido a niños en edad preescolar, mientras que 4 se aplicaban a niños entre 6 y 12 años. Estos incluían el *Physical Activity Interview* (PAI) (256), el *Computerized Activity Recall* (CAR) (257), el ACTIVITY (254) y el *Multimedia Activity Recall for Children and Adolescents* (MARCA) (258). En conjunto, mostraron una validez de constructo dispar tanto para la evaluación del comportamiento de la AF como para el GE, presentando correlaciones poco claras ($r = 0,16 - 0,88$) y resultados de fiabilidad con gran variabilidad ($ICC = 0,24 - 0,98$) (10).

Un año después, se publicó otra revisión sistemática en la que se incluyeron 29 cuestionarios específicamente diseñados para la franja de pediatría y 9 cuestionarios adaptados de poblaciones adultas. Los autores concluyeron que, aunque las medidas válidas y fiables fueron escasas, el CLASS (205) presentaba resultados prometedores (11).

En 2016 se publicó una revisión sistemática de cuestionarios de AF específicamente dirigidos a población española entre 5 y 18 años (12). Se identificaron 11 estudios que informaron sobre la fiabilidad y/o la validez de 14 cuestionarios, entre los que destacó el *Yesterday Activity Checklist* (YACH) ($r = 0,78$ e $ICC = 0,98$), dirigido a niños entre 9 y 10 años (259).

En el 2018, se actualizó la revisión sistemática publicada anteriormente por Chinapaw et al. (13). Con respecto a la validez de constructo de los cuestionarios dirigidos a edad preescolar, no se pudieron extraer conclusiones definitivas ya que los estudios fueron de baja calidad o recibieron calificaciones de evidencia negativa. Para los niños entre 6 y 12 años, se encontró que el mejor cuestionario en términos de validez de constructo era el *Godin Leisure-Time Exercise Questionnaire* (260). La

fiabilidad se evaluó en 5 cuestionarios dirigidos a preescolares, mostrando los mejores resultados el *Energy Balance Related Behaviors* (ERBs)(261). En los escolares, los cuestionarios más fiables fueron la versión china del PAQ-C (262) y el *Active Transportation to School and Work in Norway* (ATN) (212).

En resumen, la falta de estudios de alta calidad que examinen las propiedades de medición dificulta establecer recomendaciones sobre el mejor cuestionario disponible. Los estudios actuales que utilizan cuestionarios de AF en edad preescolar y escolar deben tener en cuenta que es posible que sus resultados no reflejen adecuadamente los niveles de AF ya que la mayoría de los cuestionarios carecen de la calidad suficiente (9-13).

2. Justificación

La AF es un pilar básico en la prevención de enfermedades crónicas y se relaciona con numerosos beneficios más allá de la salud física. En niños en edad preescolar existe evidencia sólida de que el volumen acumulado de AF se asocia con un riesgo reducido de aumento de peso y adiposidad (125–127) y con indicadores favorables de salud ósea (128–130). Para los niños en edad escolar, se ha demostrado que la AF regular se relaciona con una mejor resistencia cardiorrespiratoria (84,133–136), una mayor resistencia y fuerza muscular (137–140), una disminución del peso y porcentaje de grasa corporal (141–143), una mejor salud ósea (142,144–146), un menor riesgo de enfermedades cardiovasculares y metabólicas (136,147,148), una reducción de los síntomas de depresión o ansiedad (149–152) y una mejora en los indicadores de la cognición, entre ellos, la memoria, la velocidad de procesamiento, la atención y el rendimiento académico (153–155). Con respecto al impacto del CS sobre el estado de salud, la evidencia sugiere que una mayor proporción de sedentarismo está relacionada con una peor salud cardiovascular y metabólica y un mayor índice de adiposidad (156–158).

La infancia es un período clave en el desarrollo y consolidación hábitos saludables, ya que el cerebro del niño es sensible a ser influenciado por el medio ambiente en sus conductas modificables (6). Por tanto, es fundamental disponer de instrumentos válidos que permitan objetivar el nivel de AF practicado desde edades tempranas (12).

Aunque algunos investigadores consideran que el agua doblemente marcada (165), la calorimetría (170) o la observación directa (15) son los *gold standard* en la medición de los niveles de AF en los niños, muchos otros defienden que no existe un *gold standard* real (20,111,177,190–192). La evidencia científica sugiere que ningún método puede capturar todos los subcomponentes de la AF y la selección de uno u otro se debe basar en criterios objetivos en función de las indicaciones de uso, la practicidad, los

atributos cualitativos, las propiedades de medición y el objetivo de investigación (20,58,172,174).

En comparación con los métodos directos, los cuestionarios ofrecen ciertas ventajas en relación al bajo coste asociado a su distribución y a la posibilidad de recopilar información en un gran número de sujetos (58,100). Otro punto fuerte es que son capaces de identificar las dimensiones y los dominios en los que se desarrolla la AF (172,177) y no tienen riesgo de afectar a los patrones habituales de AF porque el período de referencia para medir la AF es anterior a su administración (111,186). Sin embargo, la gran cantidad de cuestionarios disponibles para la franja de pediatría y la falta de estudios de alta calidad dificultan establecer recomendaciones sobre el mejor cuestionario disponible (7-13,263).

En respuesta a la demanda mundial de medidas comparables de AF intra e inter-países, en 1996 se inició un esfuerzo para desarrollar un cuerpo de cuestionarios válidos que culminó con la integración de un Grupo de Consenso Internacional que diseñó un total de ocho versiones del IPAQ dirigidas a la población adulta. El objetivo de este grupo fue consolidar una línea de medidas indirectas que pudieran monitorizar y guiar el desarrollo de políticas enfocadas en la AF relacionada con la salud. A partir de entonces, esta familia de cuestionarios se ha ido ampliando con la creación de versiones centradas en otras poblaciones y con adaptaciones de la herramienta a diferentes países y culturas (203,264). Por ello, surge el interés de continuar el estudio de los instrumentos del grupo IPAQ, en este caso, centrándonos en los cuestionarios disponibles para niños en edad preescolar y escolar.

Por motivos de comprensión, esta tesis doctoral se compone de dos partes claramente diferenciadas.

Por un lado, se presenta una revisión sistemática sobre propiedades de medición de los cuestionarios del grupo IPAQ para niños en edad preescolar

y escolar que resume la evidencia disponible y revela la ausencia de un cuestionario dirigido a niños entre 5 y 7 años (capítulo 3).

Por otro, se describe el trabajo de investigación enmarcado en el desarrollo y estudio de propiedades de medición del *Physical Activity Questionnaire – Young Children* (PAQ-YC) dirigido a niños entre 5 y 7 años, cuyo objetivo es completar la familia de cuestionarios IPAQ disponibles (capítulo 4)

3. Cuestionarios internacionales de actividad física en edad escolar y preescolar: una revisión sistemática de propiedades de medición

3.1. Objetivos generales

- Identificar los cuestionarios del grupo International Physical Activity Questionnaire (IPAQ) dirigidos a población preescolar y escolar.
- Estudiar las propiedades de medición de los cuestionarios pertenecientes al grupo International Physical Activity Questionnaire (IPAQ) dirigidos a población preescolar y escolar

3.2. Objetivos específicos

- Calificar la calidad metodológica de los estudios que evalúan cada propiedad de medición.
- Calificar los resultados de los estudios individuales según los criterios para buenas propiedades de medición.
- Agrupar los resultados de los estudios sobre la misma propiedad de medición y calificarlos según los criterios para buenas propiedades de medición.
- Calificar la calidad de la evidencia global sobre cada propiedad de medición para cada cuestionario.

3.3. Material y métodos

Esta revisión sistemática se elaboró cumpliendo las pautas marcadas por la iniciativa COSMIN (247) y de acuerdo con las directrices de la declaración *Preferred Items for Systematic Reviews and Meta-Analyses* (PRISMA) de 2009 para revisiones sistemáticas y meta-análisis (265). El protocolo de la revisión se encuentra publicado en PROSPERO con código de registro CRD42021238246.

3.3.1. Criterios de elegibilidad

Se incluyeron los artículos que presentaran información sobre propiedades de medición de cuestionarios IPAQ en etapa escolar y preescolar.

Como criterios de inclusión, se establecieron:

- a. Estudios que incluyeran cuestionarios que registran la AF de forma exclusiva. Como excepción, se incluyeron aquellos que, además de medir la AF, registraran el CS.
- b. Estudios cuyo objetivo fuera evaluar al menos una de las propiedades de medición de un cuestionario de AF de autoinforme o informado por el cuidador.

Se excluyeron los artículos en los que:

- a. La población diana incluyera población con alguna patología u otra condición específica.
- b. La muestra del estudio no fuera representativa de la población diana a la que va dirigido el cuestionario.
- c. El cuestionario fuera administrado a un respondedor distinto al cual fue diseñado.

- d. El estudio utilizara el cuestionario como un instrumento de medición de resultados o como un instrumento de comparación en un estudio de validación.
- e. El estudio examinara la validez estructural y/o la consistencia interna de los cuestionarios que representan un modelo de medición formativo.
- f. El estudio examinara la validez de constructo como la relación entre el cuestionario y una medida diferente a la AF.
- g. El estudio examinara la capacidad de respuesta sin utilizar un instrumento de comparación validado para evaluar la capacidad del cuestionario para detectar cambios.

3.3.2. Fuentes de información y búsqueda

Se consultaron sistemáticamente las bases de datos electrónicas de MEDLINE (a través de la interfaz PubMed), Embase y Sport Discus durante los meses de octubre de 2020 a enero de 2021. Se descartó acudir a fuentes de información adicionales como editoriales, actas de congresos, informes, etc. porque la probabilidad de encontrar artículos sobre propiedades de medición en este tipo de fuentes es mínima (247). Se completó la búsqueda con una revisión manual de las referencias en los artículos incluidos y una búsqueda simple en internet para localizar artículos no indexados en revistas de impacto.

La estrategia de búsqueda para las diferentes bases de datos consistió en una colección completa de términos para los cuatro elementos clave del objetivo de la revisión: constructo, población, tipo de instrumento y propiedades de medición. En Sport Discus, los términos clave se diseñaron tomando como referencia los “bloques de búsqueda” desarrollados por la biblioteca de la Universidad Vrije de Amsterdam (266). Para las bases de datos MEDLINE y Embase, se utilizó el filtro de búsqueda para estudios sobre propiedades de

medición desarrollado por la Universidad de Oxford en el año 2009 (sensibilidad 97,4% y precisión 4,4%) (267). El filtro incluyó términos que identifican propiedades de medición irrelevantes en los cuestionarios de AF, como la consistencia interna o la validez estructural. Se decidió mantener esta estrategia y revisar manualmente todos los estudios incluidos por la posibilidad de que algunos autores emplearan una terminología inadecuada. Para definir qué propiedad de medición se analizaba en cada estudio se siguieron las directrices de la taxonomía COSMIN (Tabla 2) (239). La estrategia de búsqueda detallada para cada base de datos se presenta en la Tabla 6.

Tabla 6. Estrategia de búsqueda en PubMed, EMBASE y Sport Discus

PubMed	EMBASE	Sport Discus
#1 AND #2 AND #3 AND #4 NOT #5	#1 AND #2 AND #3 AND #4	#1 AND #2 AND #3
<p>#1 construct search "exercise"[Majr] OR "motor activity"[Majr] OR "physical exertion*" [ti] OR "physical effort*" [ti] OR "Physical Activit*" [ti] OR "Exercis*" [ti]</p> <p>#2 population search "Child"[Mesh] OR "Child, Preschool"[Mesh]</p> <p>#3 instrument "Surveys and Questionnaires"[Majr] OR "Health Surveys"[Majr] OR "Health Care Surveys"[Majr] OR "Patient Reported Outcome Measures"[Majr] OR "Self Report"[Majr] OR "question*" [ti] OR "survey*" [ti] OR "report*" [ti] OR "self-report*" [ti] OR "self report*" [ti] OR "assess*" [ti] OR "self-assess*" [ti] OR "self-administ*" OR "measure*" [ti] OR "recall" [ti]</p> <p>#4 measurement properties (instrumentation[sh] OR methods[sh] OR "Validation Study"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment, health care"[MeSH] OR "outcome assessment" [tiab] OR "outcome measure*" [tw] OR "observer variation"[MeSH] OR "observer variation" [tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant</p>	<p>#1 construct search 'physical activity'/mj OR 'motor activity'/mj OR 'exercise'/mj OR 'energy expenditure'/mj OR 'energy metabolism'/mj OR 'physical exertion*':ti OR 'physical effort*':ti OR 'Physical Activit*':ti OR 'Exercis*':ti</p> <p>#2 population search 'child'/de OR 'preschool child'/de OR 'school child'/de</p> <p>#3 instrument 'questionnaire'/mj OR 'self report'/mj OR 'clinical assessment tool'/mj OR 'rating scale'/mj OR 'recall'/mj OR 'question*':ti OR 'survey*':ti OR 'report*':ti OR 'self-report*':ti OR 'self report*':ti OR 'assess*':ti OR 'self-assess*':ti OR 'self-administ*' OR 'measure*':ti OR 'recall':ti</p> <p>#4 measurement properties 'intermethod comparison'/exp OR 'data collection method'/exp OR 'validation study'/exp OR 'feasibility study'/exp OR 'pilot study'/exp OR 'psychometry'/exp OR 'reproducibility'/exp OR reproducib*:ab,ti OR 'audit':ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR 'observer variation'/exp OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'validity'/exp OR reliab*:ab,ti OR valid*:ab,ti OR 'coefficient':ab,ti OR 'internal consistency':ab,ti OR</p>	<p>#1 construct search (DE "EXERCISE") OR (DE "PHYSICAL activity") OR (DE "MOTOR ability") OR (TI "physical exertion*") OR (TI "physical effort*") OR (TI "Physical Activit*") OR (TI "Exercis*")</p> <p>#2 population search (DE "CHILDREN") OR (DE "SCHOOL children")</p> <p>#3 instrument (DE "PHYSICAL activity measurement") OR (DE "SELF-evaluation") OR (TI "question*") OR (TI "survey*") OR (TI "report*") OR (TI "self-report*") OR (TI "self report*") OR (TI "assess*") OR (TI "self-assess*") OR (TI "self-administ*") OR (TI "meaure*") OR (TI "recall")</p>

<p>analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor</p>	<p>(cronbach*:ab,ti AND ('alpha':ab,ti OR 'alphas':ab,ti)) OR 'item correlation':ab,ti OR 'item correlations':ab,ti OR 'item selection':ab,ti OR 'item selections':ab,ti OR 'item reduction':ab,ti OR 'item reductions':ab,ti OR 'agreement':ab,ti OR 'precision':ab,ti OR 'imprecision':ab,ti OR 'precise values':ab,ti OR 'test-retest':ab,ti OR ('test':ab,ti AND 'retest':ab,ti) OR (reliab*:ab,ti AND ('test':ab,ti OR 'retest':ab,ti)) OR 'stability':ab,ti OR 'interrater':ab,ti OR 'inter-rater':ab,ti OR 'intrarater':ab,ti OR 'intra-rater':ab,ti OR 'intertester':ab,ti OR 'inter-tester':ab,ti OR 'intratester':ab,ti OR 'intra-tester':ab,ti OR 'interobserver':ab,ti OR 'inter-observer':ab,ti OR 'intraobserver':ab,ti OR 'intra-observer':ab,ti OR 'intertechnician':ab,ti OR 'inter-technician':ab,ti OR 'intratechnician':ab,ti OR 'intra-technician':ab,ti OR 'interexaminer':ab,ti OR 'inter-examiner':ab,ti OR 'intraexaminer':ab,ti OR 'intra-examiner':ab,ti OR 'interassay':ab,ti OR 'inter-assay':ab,ti OR 'intraassay':ab,ti OR 'intra-assay':ab,ti OR 'interindividual':ab,ti OR 'inter-individual':ab,ti OR 'intraindividual':ab,ti OR 'intra-individual':ab,ti OR 'interparticipant':ab,ti OR 'inter-participant':ab,ti OR 'intraparticipant':ab,ti OR 'intra-participant':ab,ti OR 'kappa':ab,ti OR 'kappas':ab,ti OR 'coefficient of variation':ab,ti OR repeatab*:ab,ti OR (replicab*:ab,ti OR 'repeated':ab,ti AND ('measure':ab,ti OR 'measures':ab,ti OR 'findings':ab,ti OR 'result':ab,ti OR 'results':ab,ti OR 'test':ab,ti OR 'tests':ab,ti)) OR generaliza*:ab,ti OR generalisa*:ab,ti OR 'concordance':ab,ti OR ('intraclass':ab,ti AND correlation*:ab,ti) OR 'discriminative':ab,ti OR 'known group':ab,ti OR 'factor analysis':ab,ti OR 'factor analyses':ab,ti OR 'factor structure':ab,ti OR 'factor</p>
--	--

<p>analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab])) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])</p>	<p>structures':ab,ti OR 'dimensionality':ab,ti OR subscale*:ab,ti OR 'multitrait scaling analysis':ab,ti OR 'multitrait scaling analyses':ab,ti OR 'item discriminant':ab,ti OR 'interscale correlation':ab,ti OR 'interscale correlations':ab,ti OR ('error':ab,ti OR 'errors':ab,ti AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR 'accuracy':ab,ti OR 'accurate':ab,ti OR 'precision':ab,ti OR 'mean':ab,ti)) OR 'individual variability':ab,ti OR 'interval variability':ab,ti OR 'rate variability':ab,ti OR 'variability analysis':ab,ti OR ('uncertainty':ab,ti AND ('measurement':ab,ti OR 'measuring':ab,ti)) OR 'standard error of measurement':ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR ('limit':ab,ti AND 'detection':ab,ti) OR 'minimal detectable concentration':ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND ('real':ab,ti OR 'detectable':ab,ti) AND ('change':ab,ti OR 'difference':ab,ti)) OR 'meaningful change':ab,ti OR 'minimal important change':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important change':ab,ti OR 'minimally important difference':ab,ti OR 'minimal detectable change':ab,ti OR 'minimal detectable difference':ab,ti OR 'minimally detectable change':ab,ti OR 'minimally detectable difference':ab,ti OR 'minimal real change':ab,ti OR 'minimal real difference':ab,ti OR 'minimally real change':ab,ti OR 'minimally real difference':ab,ti OR 'ceiling effect':ab,ti OR 'floor effect':ab,ti OR 'item response model':ab,ti OR 'irt':ab,ti OR 'rasch':ab,ti OR 'differential item functioning':ab,ti OR 'dif':ab,ti OR 'computer adaptive testing':ab,ti OR 'item bank':ab,ti OR 'cross-cultural equivalence':ab,ti</p>
<p>#5 exclusion filter ("address"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lecture"[Publication Type] OR "legal case"[Publication Type] OR</p>	

<p>“legislation”[Publication Type] OR “letter”[Publication Type] OR “news”[Publication Type] OR “newspaper article”[Publication Type] OR “patient education handout”[Publication Type] OR “popular work”[Publication Type] OR “congress”[Publication Type] OR “consensus development conference”[Publication Type] OR “consensus development conference, nih”[Publication Type] OR “practice guideline”[Publication Type]) NOT (“animals”[MeSH Terms] NOT “humans”[MeSH Terms])</p>		
---	--	--

3.3.3. Selección de los artículos

Se evaluó la elegibilidad de todos los artículos en base a los criterios de selección definidos. Dos revisores (CB y AM) examinaron de forma independiente el título y el resumen y leyeron detenidamente el texto completo si existía ambigüedad en cuanto a su inclusión. En caso de desacuerdo y/o discrepancia, el artículo fue analizado por un asesor independiente (MA) para decidir su inclusión final.

3.3.4. Evaluación de la calidad metodológica de los estudios

Dos revisores distintos a los anteriores (MG y MA) evaluaron de forma independiente la calidad metodológica de cada estudio utilizando el Checklist de Riesgo de Sesgo de COSMIN actualizado en el 2018 (250). En caso de desacuerdo y/o discrepancia, el equipo investigador decidió por consenso la calificación asignada.

El Checklist de Riesgo de Sesgo de COSMIN se utilizó como una herramienta modular, es decir, completando solo la información correspondiente a las propiedades de medición que se evaluaban en cada artículo (Anexo 8.1). Si en un artículo se describían varios estudios sobre diferentes propiedades de medición, la calidad de cada uno se evaluaba por separado utilizando el checklist de COSMIN correspondiente. Por otro lado, si se informaba la misma propiedad de medición para varios subgrupos o con diferentes metodologías en un mismo artículo, los resultados se presentaron de forma separada como estudios independientes.

Cada estudio se identificó con una puntuación del 1 al 4 según si se había calificado como “inadecuado”, “dudoso”, “adecuado” o “excelente”, respectivamente. Para determinar la puntuación general de la calidad de cada estudio individual para una propiedad de medición, se tomó la calificación más baja de cualquier estándar del checklist.

No se pudo evaluar la validez de contenido en ninguno de los casos ya que se disponía de muy poca información sobre el proceso de desarrollo de los cuestionarios, así como sobre la evaluación de la relevancia, exhaustividad y comprensibilidad de los ítems. En cambio, se incluyó una descripción de cada cuestionario analizado que se presenta en el apartado de resultados.

Tampoco se pudo evaluar la calidad metodológica de los estudios de validez transcultural ya que los estándares incluidos en el Checklist de Riesgo de Sesgo de COSMIN son propios de cuestionarios diseñados bajo un modelo reflexivo (250). Los métodos estadísticos propuestos en el checklist responden a teorías de medición como la Teoría Clásica de Medición o la Teoría de Respuesta al Ítem (228). Sin embargo, en el apartado de resultados se incluyó una descripción de las características generales de los estudios de validez transcultural en cuanto al país e idioma de validación, la edad, el tamaño de la muestra y los resultados de la puntuación.

Para evaluar la calidad metodológica de los estudios de fiabilidad y error de medición se siguieron los estándares descritos previamente por Chinapaw et al. sobre el intervalo de tiempo óptimo entre el test y el retest: (a) entre 1 día y 3 meses para los cuestionarios con período de recuerdo de “una semana habitual, (b) entre 1 día y 2 semanas para los cuestionarios con período de recuerdo de “la última semana” y (c) entre 1 día y 1 semana para los cuestionarios con período de recuerdo del “último día”.

Para evaluar la calidad metodológica de los estudios de validez de constructo se tuvieron en cuenta los niveles de evidencia publicados por Hidding et al. para cuestionarios de AF (Tabla 4) (13) y para cuestionarios sobre CS (Tabla 5) (248). La calificación del estándar 4 (*“¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?”*) se definió como “excelente” si la correlación entre el constructo del cuestionario y el instrumento comparador se clasificaban como nivel de evidencia 1. Se asignó un valor de “dudoso” si se comparaban los resultados del cuestionario con instrumentos

capaces de aportar un nivel de evidencia 2 o 3. Para el resto de los casos, se consideró que el estándar 4 era “inadecuado”. Con respecto a la suficiencia de las propiedades de medición del instrumento comparador, se realizó una búsqueda bibliográfica si los autores no proporcionaban esta información en el estudio.

Por último, no se pudo evaluar la capacidad de respuesta ya que no se encontraron estudios que evaluaran esta propiedad de medición en ninguno de los cuestionarios incluidos.

3.3.5. Aplicación de los criterios para buenas propiedades de medición

Dos revisores independientes (MG y MA) evaluaron el resultado de cada estudio según los criterios para buenas propiedades de medición. En caso de desacuerdo y/o discrepancia, el equipo investigador decidió por consenso la calificación asignada. Cada resultado se puntuó como suficiente (+), insuficiente (-) o indeterminado (?).

Los resultados de fiabilidad test-retest se consideraron suficientes (+) si el Coeficiente de Correlación Interclase (ICC) o el índice Kappa eran iguales o superiores a 0,7 (242,247). En caso de que la fiabilidad se presentase a través de índices de correlación de Pearson o Spearman se estableció que era suficiente (+) si alcanzaba un valor mínimo de 0,8 (201). Para aquellos estudios en los que se presentaba la fiabilidad test-retest tanto con el ICC o Kappa, como con el coeficiente de Pearson o Spearman, se evaluaron únicamente los resultados del ICC o Kappa por ser los métodos estadísticos preferidos (250).

Los resultados del error de medición se consideran aceptables (+) cuando el cambio detectable más pequeño (SDC), el límite del acuerdo (LoA) o el porcentaje de acuerdo (PoA) es menor que el cambio mínimo importante (247). Desafortunadamente, no se pudo presentar una calificación de

evidencia para el error de medición, ya que según nuestro conocimiento, no hay información disponible sobre el cambio mínimo importante de los cuestionarios incluidos. Por tanto, todos los resultados fueron calificados como indeterminados (?) y se finalizó el análisis de esta propiedad de medición.

De acuerdo con las recomendaciones de la guía COSMIN para evaluar los resultados de los estudios sobre validez de constructo, se establecieron hipótesis a priori respecto a la correlación esperada con otros instrumentos. Los resultados se subdividieron por niveles de evidencia según los criterios publicados por Hidding et al. en 2018 para cuestionarios de AF (ver Tabla 4) (13) y para cuestionarios sobre CS (ver Tabla 5) (248).

3.3.6. Resumen de la evidencia por propiedad de medición

Para el análisis y síntesis de los datos se utilizó el programa *MedCalc Statistical Software 19.7.2*. El efecto de medida de resultado fue la estimación del coeficiente de correlación o ICC y su intervalo de confianza al 95% (IC 95%) utilizando el modelo de efectos aleatorios ponderando por el inverso del tamaño de muestra de cada uno de los estudios.

En primer lugar, se agruparon los resultados de todos los estudios disponibles para cada propiedad de medición y se evaluó si eran consistentes o inconsistentes con el test de heterogeneidad Q-Cochrane y calculando el coeficiente de inconsistencia (I^2). De acuerdo con la clasificación de Higgins y Thompson, se estableció que los resultados eran consistentes si el valor de I^2 era menor o igual al 25%. La inconsistencia se clasificó como baja, moderada o alta según si el valor de I^2 se encontraba entre 25 – 50%, entre 50 – 75% y más 75%, respectivamente (268,269).

De acuerdo con las recomendaciones de la guía COSMIN (247), si los resultados eran consistentes, se compararon de nuevo con los criterios para

buenas propiedades de medición para determinar si, en general, la propiedad de medición del cuestionario era suficiente (+), insuficiente (-) o indeterminada (?). Si los resultados eran inconsistentes, se decidió en cada caso la estrategia más apropiada. En primer lugar, se exploraron las posibles explicaciones a la inconsistencia y, si estaba justificado, se realizaron los análisis por subgrupos pertinentes para presentar los resultados de forma separada. Si el análisis por subgrupos no resolvía la inconsistencia y no se encontraron explicaciones adicionales, el resultado se calificó igualmente según los criterios para buenas propiedades de medición y, posteriormente, la calidad de la evidencia se degradó debido a la inconsistencia.

3.3.7. Calificación de la calidad de la evidencia

En base a las recomendaciones de la guía COMSIN, se estableció que la calidad de la evidencia era alta si “se consideraba que la verdadera propiedad de medición se encontraba cerca de la estimación dada”; moderada si “era probable que la propiedad de medición real estuviera cerca de la estimación dada, pero existía la posibilidad de que fuera sustancialmente diferente”; baja si “la confianza en la estimación era limitada y se reconocía que la propiedad de medición real podía ser sustancialmente diferente”; y muy baja si “la confianza en la estimación era muy pobre y se reconocía que era probable que la propiedad de medición real fuera sustancialmente diferente de la estimación dada” (247). Adicionalmente, para la validez de constructo, la calidad de la evidencia se clasificó por niveles 1, 2 o 3 según los criterios publicados por Hidding et al. en 2018 para cuestionarios de AF (Tabla 4) (13) y para cuestionarios sobre CS (ver Tabla 5) (248). El nivel 1 proporciona evidencia sólida, el nivel 2 es indicativo de evidencia moderada y el nivel 3 ofrece evidencia débil.

La clasificación de la calidad de la evidencia se basó en el enfoque modificado del *Grading of Recommendations Assessment, Development, and Evaluation*

(*GRADE*) propuesto por la guía COSMIN para revisiones sistemáticas de propiedades de medición (247). El *GRADE* modificado se utilizó como una herramienta de degradación de la calidad en base a la existencia de riesgo de sesgo, inconsistencia y/o imprecisión. Aunque también se recomienda rebajar la calidad por resultados indirectos, no se consideró necesario en este estudio ya que los artículos que aplicaban el cuestionario a una muestra diferente a la población diana fueron excluidos.

En primer lugar, la calidad de la evidencia se degradó un nivel (de evidencia alta a moderada) si existía riesgo grave de sesgo; dos niveles (de evidencia alta a baja) si el riesgo de sesgo era muy grave y tres niveles (de evidencia alta a muy baja) si el riesgo de sesgo era extremo. Se consideró que no existía riesgo de sesgo si había al menos un estudio de calidad metodológica “excelente” o si la mayoría de los estudios presentaban puntuaciones de “adecuada” calidad, según lo evaluado con el Checklist de Riesgo de Sesgo de COSMIN (250). Se estableció que el riesgo de sesgo era grave si la mayoría de los estudios presentaban “dudosa” calidad metodológica o si había un único estudio calificado como “adecuado”. Se definió un riesgo de sesgo muy grave si la mayor parte de los estudios mostraban una calidad “inadecuada” o si solo se disponía de un estudio de calidad “dudosa”. Por último, se decidió que existía riesgo de sesgo extremo si solo había disponible un estudio y su calidad metodológica era “inadecuada” (247).

En segundo lugar, se degradó la calidad de la evidencia por inconsistencia en base a los criterios definidos Higgins y Thompson (268,269). Se consideró rebajar un nivel si el valor de I^2 era del 50-75% (inconsistencia moderada) y dos niveles si superaba el 75% (inconsistencia alta).

Por último, la calidad de la evidencia se rebajó por imprecisión en un nivel si el tamaño total de la muestra de los estudios agrupados era inferior a 100 individuos y en dos niveles si era inferior a 50.

3.4. Resultados

La búsqueda en las bases de datos electrónicas de MEDLINE (a través de la interfaz PubMed), Embase y Sport Discus ofreció un total de 2.090 resultados después de eliminar los duplicados. Se completó con una revisión manual y una búsqueda simple en internet que arrojó 8 artículos no indexados en revistas de impacto susceptibles de ser incluidos en el estudio. Tras leer el título y resumen, se descartaron 1.926 artículos, quedando un total de 164 registros para ser examinados por texto completo. Finalmente, se incluyeron un total de 21 artículos en el análisis cualitativo y 17 artículos en el análisis cuantitativo. En la Figura 17 se muestra el diagrama de flujo detallado.

De los cuestionarios dirigidos a edad preescolar, se incluyó 1 artículo que presenta el estudio cualitativo del desarrollo del EY-PAQ y muestra los resultados del estudio de fiabilidad y validez de constructo. Para edad escolar, se incluyeron un total de 16 artículos, de los cuales 2 estudios se corresponden con los estudios de validación original del PAQ-C, 18 estudios se incluyen como parte del estudio cualitativo de traducción y adaptación transcultural, 14 estudios (correspondientes a 12 artículos) reportan resultados de fiabilidad, 5 estudios (correspondientes a 3 artículos) muestran el error de medición entre el test y el retest y 16 estudios (correspondientes a 9 artículos) evalúan la validez de constructo por comparación de métodos.

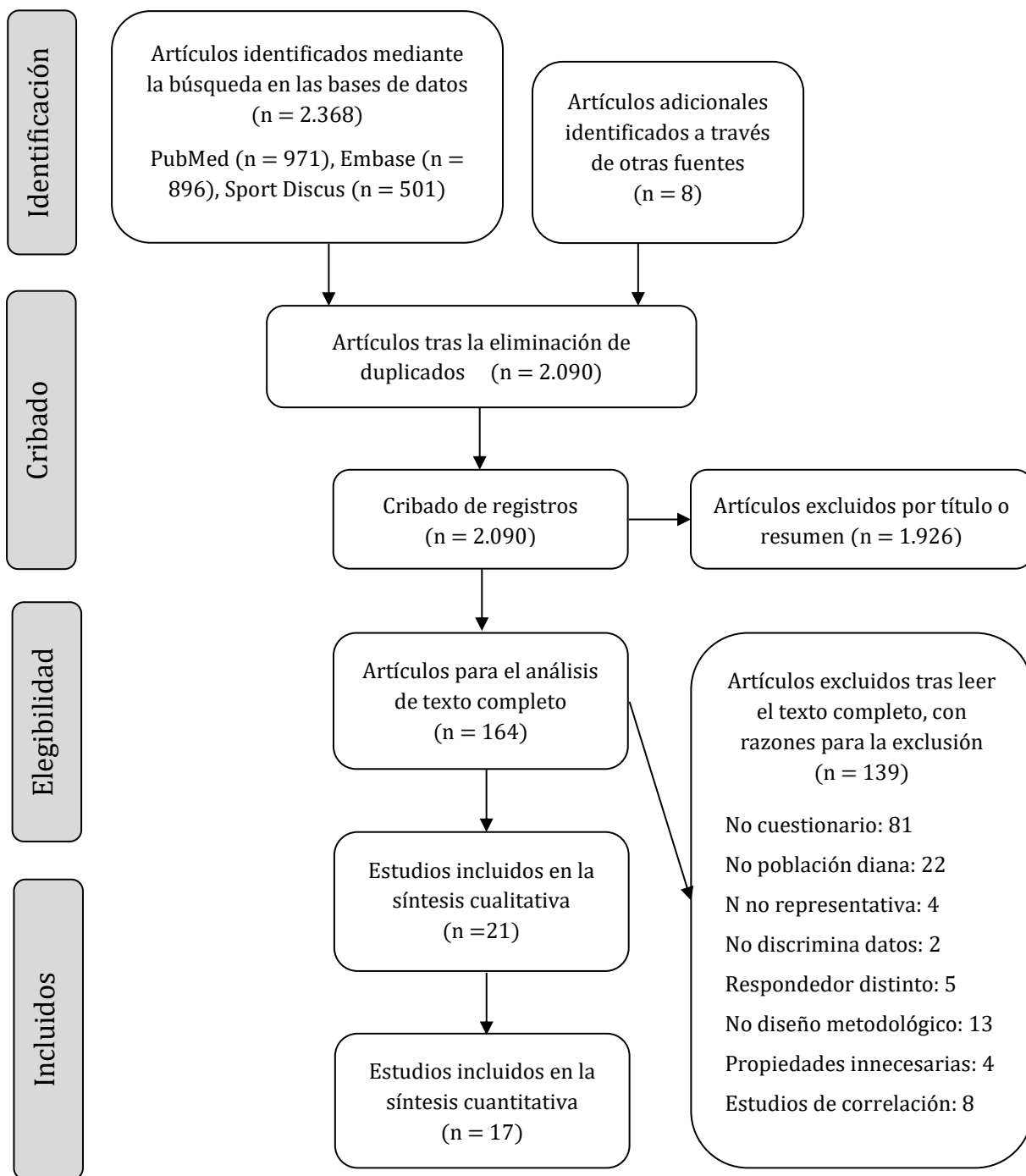


Figura 17. Diagrama de flujo PRISMA para la selección de los artículos incluidos.

3.4.1. Validez de contenido

Tal y como se ha adelantado en el apartado de material y métodos, no se pudo evaluar ni la calidad metodológica ni los criterios para una adecuada validez de contenido de los cuestionarios incluidos, ya que en ninguno de los casos se disponía de suficiente información sobre el proceso de desarrollo. A continuación, se proporciona una descripción cualitativa de las características relevantes de cada uno de ellos, descritas por los autores de las validaciones originales.

Por un lado, el EY-PAQ es un cuestionario informado por los padres que cuantifica los niveles de AFMV y CS practicados en una semana habitual del último mes en niños menores de 5 años. Fue publicado en 2016 por Bingham et al. en Reino Unido y está disponible en los idiomas urdu e inglés. Las actividades que recoge como AFMV se engloban dentro de los juegos activos en el hogar, los juegos activos en el jardín, los desplazamientos activos, los juegos que provocan sudoración y aumento de la frecuencia respiratoria, los juegos en el parque o el patio del recreo y los juegos en instalaciones interiores. Las actividades que se incluyen como CS son colorear, dibujar, hacer manualidades, sentarse a jugar con juguetes, ver la televisión, jugar al ordenador, sentarse a escuchar música o cantar, leer o escuchar un cuento y desplazarse de forma pasiva (en un cochecito, en una silla de paseo, en brazos, en coche o en transporte público). La frecuencia se recoge en una casilla que pregunta por el número de días a la semana que se realiza cada actividad. La duración se presenta como intervalos de tiempo que van de menos de 15 minutos, de 16 a 30 minutos, de 31 a 60 minutos y más de una hora al día (224) (Anexo 8.2).

Para calcular la puntuación del EY-PAQ se debe seguir un proceso de tres pasos. En primer lugar, los minutos diarios de AFMV y CS se multiplican por la frecuencia con la que se produce cada actividad. Con respecto al componente de duración, se utiliza un enfoque pragmático al escoger el límite

inferior del intervalo si se trata de una AFMV y el rango superior si la actividad encaja dentro del CS, a excepción de que la respuesta sea texto libre. En segundo lugar, se suman los minutos de todas las actividades que conforman la AFMV y el CS y se dividen por 7 para obtener una estimación de la duración diaria. En tercer lugar, se dividen los minutos de AFMV y CS entre 814 minutos (14 horas) para convertirlos en las proporciones de tiempo de vigilia (224).

Por otro lado, el PAQ-C es un cuestionario autoinformado que valora la participación en AFMV en período escolar durante los últimos 7 días en niños de 8 a 14 años. Los estudios de validación original fueron publicados en 1997 por Crocker y Kowalski (208,253) en Canadá, en lengua inglesa. El cuestionario que presentan consta de 10 preguntas de las cuales solo las 9 primeras se utilizan como indicadores de los niveles de AFMV. Los ítems del 1 al 7 evalúan las actividades realizadas durante el tiempo libre (formato de pregunta tipo checklist), durante las clases de Educación Física, en el recreo, en la pausa a mediodía, en el período extraescolar, por las tardes y el fin de semana. Los dos últimos ítems preguntan por la frecuencia en la participación de AFMV de manera general (ítem 8) y para cada día de la semana (ítem 9). La pregunta 10 valora si los últimos 7 días pueden considerarse representativos de una semana habitual o si ha habido alguna incidencia que haya impedido el desempeño de las actividades usuales. Cabe destacar que en ningún ítem se pide información sobre la duración y únicamente se registra la frecuencia (208,253) (Anexo 8.3).

Para calcular la puntuación del PAQ-C, se asigna a cada ítem una calificación en una escala Likert de 1 a 5 (excepto para el ítem 10) y se presenta la media para obtener el resultado final, donde las puntuaciones más altas indican mayores niveles de actividad. En las preguntas 1 y 9, se calcula la media de todas las actividades del checklist y de los días de la semana, respectivamente (208,253).

3.4.2. Validez transcultural

Con respecto al EY-PAQ, no se disponía de ningún estudio de validez transcultural a otro idioma y/o población culturalmente diferente a la original (224).

En relación al PAQ-C, se han descrito procesos cualitativos de traducción y/o adaptación transcultural al inglés de EE.UU. (270), al persa (271), al español de Colombia (272) y de España (273,274), al croata (275), al inglés de Reino Unido (276), al holandés de Bélgica (277), al portugués de Brasil (278), al malayo (279), al chino cantonés (262), al italiano (280), al afrikáans (281), al árabe (282), al turco (283), al japonés (236), al gujaratí (284) y al tailandés (285). Tal y como se ha explicado en el apartado de material y métodos, no se evaluó ni la calidad metodológica ni los criterios para una adecuada validez transcultural en ninguno de los estudios debido a que los estándares incluidos en el Checklist de Riesgo de Sesgo de COSMIN no se ajustan a cuestionarios basados en modelos formativos (250).

En la Tabla 7 se presentan las características generales de los estudios originales del EY-PAQ (224) y del PAQ-C (208,253), así como de los artículos incluidos en el análisis narrativo de la validez transcultural del PAQ-C. Para cada estudio se presenta el país e idioma de validación, la edad de la muestra (rango, media y desviación típica), el tamaño de la muestra (total y por sexo) y los resultados de la puntuación (total y por sexos).

Tabla 7. Características generales de los artículos de validación original y de validez transcultural

	Estudio	País	Idioma	Edad muestra	Edad media (SD)	N total	Puntuación total	N niños	% niños	Puntuación niños	N niñas	% niñas	Puntuación niñas
EY-PAQ	Bingham 2016 (224)	Reino Unido: VO	inglés urdu	< 5	3,20 (0,8)	156 40	AFMV: 21,2 (11) CS: 47,0 (13,60)	99	50,5	- -	97	49,5	- -
PAQ-C	Crocker 1997 (253)	Canadá: VO	inglés	9 - 15	-	215	3,24 (0,72)	125	58,14	3,44 (0,68)	90	41,86	2,96 (0,69)
	Kowalski 1997 (208)	Canadá: VO	inglés	8 - 13	11,06 (1,46)	84	3,23 (0,78)	36	42,86	-	48	57,14	-
				9 - 14	11,30 (0,39)	97	3,35 (0,68)	41	42,27	-	56	57,73	-
	Moore 2007 (270)	EE.UU.	inglés	8 - 9	8,70 (0,57)	414	3,36 (0,80) AE 3,37 (0,69) AA	213	51,45	-	201	48,55	-
	Faghihimani 2010 (271)	Irán	persa	7 - 18	-	82	-	20	24,39	-	62	75,61	-
	Herazo-Beltrán 2012 (272)	Colombia	español	9 - 11	9,35 (1,05)	100	1,90 (0,30)	52	52,00	2,20 (0,4)	48	48,00	1,80 (0,3)
	Vidaković-Samaržija 2013 (275)	Croacia	croata	10	-	56	3,08 (0,68)	23	41,07	-	23	41,07	-
	Thomas 2014 (276)	Reino Unido	inglés	9 - 11	9,93 (0,80)	336	3,49 (0,68)	168	50,00	-	168	50,00	-
	Bervoets 2014 (277)	Bélgica	holandés	5 - 12	8,90 (1,70)	192	3,05 (0,89)	91	47,40	-	101	52,60	-
	Guedes 2015 (278)	Brasil	portugués	8 - 13	11,12 (1,38)	232	2,84 (0,57)	108	46,55	3,21 (0,61)	124	53,45	2,62 (0,53)
	Zaki 2015 (279)	Malasia	malayo	10 - 17	-	73	-	25	34,25	2,53 (0,56)	48	65,75	2,36 (0,62)
	Benítez-Porres 2016 (273)	España	español	8 - 13	10,98 (1,17)	83	3,24 (0,64)	46	55,42	3,22 (0,70)	37	44,58	3,26 (0,56)
	Wang 2016 (262)	China	cantonés	8 - 13	10,5 (1,1)	742	2,62 (0,68)	412	55,53	2,67 (0,70)	330	44,47	2,56 (0,66)
Gobbi 2016 (280)	Italia	italiano	8 - 11	9,30 (0,60)	1116	3,05 (0,67)	584	52,33	3,12 (0,66)	532	47,67	2,93 (0,66)	

Malan 2017 (281)	Sudáfrica	afrikáans	7 - 9	-	58	3,14 (0,47)	30	51,72	-	28	48,28	-
Manchola-Glez. 2017 (274)	España	español	8 - 14	11,60 (1,40)	72	2,28 (0,54)	38	52,78	2,39 (0,50)	34	47,22	2,15 (0,56)
Ben Jemaa 2018 (282)	Túnez	árabe	8 - 11	9,34 (0,94)	40	2,55 (0,67)	21	52,50	2,82 (0,71)	19	47,50	2,24 (0,47)
Erdim 2019 (283)	Turquía	turco	9 - 14	11,00 (1,34)	784	3,16 (0,73)	402	51,28	3,27 (0,74)	382	48,72	3,04 (0,69)
Isa 2019 (236)	Japón	japonés	9 - 12	10,90 (0,90)	184	2,76 (0,69)	104	56,52	2,84 (0,70)	80	43,48	2,65 (0,68)
Zishan 2019 (284)	India	guyaratí	8 - 14	12,75 (1,29)	20	3,08 (0,77)	12	60,00	-	8	40,00	-
Pratanaphon 2020 (285)	Tailandia	tailandés	8 - 14	10,32 (1,87)	115	2,67 (0,61)	57	49,57	2,80 (0,56)	58	50,43	2,53 (0,63)

EY-PAQ: Early Years - Physical Activity Questionnaire, PAQ-C: Physical Activity Questionnaire – Older Children, SD: desviación estándar, N: tamaño de la muestra, VO: versión original, AFMV: actividad física moderada-vigorosa, CS: comportamiento sedentario, AE: europeo-americano, AA: afroamericano

3.4.3. Fiabilidad

En la Tabla 8 se muestran las características de los estudios que evalúan la fiabilidad del EY-PAQ y el PAQ-C con respecto al intervalo entre el test y el retest, la calificación desglosada y general de la calidad metodológica del estudio, el tamaño de muestra, el resultado y la calificación del resultado según los criterios establecidos.

Tabla 8. Estudios de fiabilidad. Evaluación de la calidad metodológica y calificación del resultado

	Estudio	Intervalo test-retest	Calidad metodológica				N	Resultado	Calificación
			RD	AE	O	G			
EY-PAQ	Bingham 2016 (224)	5-12 días	2	4	4	2	104	ICC = 0,35 AFMV ICC = 0,47 CS	(-)
PAQ-C	Crocker 1997 (253)	7 días	2	3	4	2	43	ICC = 0,75	(+)
	Crocker 1997 (253)	7 días	2	3	4	2	41	ICC = 0,82	(+)
	Herazo-Beltrán 2012 (272)	7 días	2	3	4	2	100	ICC = 0,61	(-)
	Storey 2012 (232)	8 días	2	3	4	2	323	ICC = 0,79	(+)
	Guedes 2015 (278)	7 días	2	3	4	2	232	ICC = 0,74	(+)
	Benítez-Porres 2016 (273)	< 24 horas	1	4	4	1	83	ICC = 0,96	(+)
	Wang 2016 (262)	7-10 días	2	4	4	2	92	ICC = 0,82	(+)
	Manchola-Glez. 2017 (274)	2 horas	1	3	1	1	72	ICC = 0,85	(+)
	Manchola-Glez. 2017 (274)	7 días	2	3	4	2	72	ICC = 0,80	(+)
	Podnar 2017 (230)	3 semanas	1	3	4	1	46	ICC = 0,75	(+)
	Erdim 2019 (283)	10 días	2	3	4	2	71	ICC = 0,91	(+)
	Isa 2019 (236)	2 meses	1	4	4	1	154	ICC = 0,83	(+)
	Zishan 2019 (284)	1 día	2	3	4	2	20	ICC = 0,94	(+)
	Pratanaphon 2020 (285)	2 semanas	2	3	4	2	115	ICC = 0,67	(-)

EY-PAQ: Early Years – Physical Activity Questionnaire, PAQ-C: Physical Activity Questionnaire – Older Children, RD: requisitos de diseño, AE: análisis estadístico, O: otros, G: general, 4: “excelente”, 3: “adecuado”, 2: “dudoso”, 1: “inadecuado”, N: tamaño de muestra, ICC: índice de concordancia intraobservador, AFMV: actividad física moderada-vigorosa, CS: comportamiento sedentario

Para el EY-PAQ, al no haberse publicado ningún otro estudio sobre su fiabilidad, se determinó que era insuficiente y la calidad de la evidencia se calificó como muy baja.

Para el PAQ-C, se agruparon todos los estudios de fiabilidad disponibles (14 estudios correspondientes a 12 artículos) (Tabla 9 y Figura 18). Se obtuvo un ICC total de 0,82 (IC 95% [0,76 - 0,87]) con un porcentaje de inconsistencia del 88,25% (IC 95% [82,03 - 92,32]). Se estableció que existía inconsistencia alta según los criterios de Higgins y Thompson (268,269) y se procedió a realizar el análisis por subgrupo de los estudios de fiabilidad que habían obtenido al menos una puntuación de “dudosa” en la calificación de la calidad metodológica (Tabla 10 y Figura 19). El resultado de este meta-análisis (10 estudios correspondientes a 9 artículos) presentó un ICC total de 0,79 (IC 95% [0,73 - 0,84]) con un porcentaje de inconsistencia del 80,10% (IC 95% [64,19 - 88,94]), calificado también como inconsistencia alta (268,269). Se concluyó que la síntesis por subgrupo no resolvía la inconsistencia y no se encontraron explicaciones adicionales que justificasen nuevos subanálisis. Los resultados se puntuaron como suficientes (+) en ambos casos, aunque la calidad de la evidencia se degradó en tres niveles (calidad de la evidencia muy baja) por existencia de riesgo de sesgo grave e inconsistencia alta.

Tabla 9. Meta-análisis de todos los estudios de fiabilidad del PAQ-C

Estudio	N	ICC	IC 95%	Efectos aleatorios
Crocker 1997 (253)	43	0,75	[0,58 - 0,86]	6,46
Crocker 1997 (253)	41	0,82	[0,69 - 0,90]	6,38
Herazo-Beltrán 2012 (272)	100	0,61	[0,47 - 0,72]	7,54
Storey 2012 (232)	323	0,79	[0,75 - 0,83]	8,21
Guedes 2015 (278)	232	0,74	[0,68 - 0,79]	8,08
Benítez-Porres 2016 (273)	83	0,96	[0,94 - 0,97]	7,35
Wang 2016 (262)	92	0,82	[0,74 - 0,88]	7,46

Manchola-Glez. 2017 (274)	72	0,85	[0,77 - 0,90]	7,20
Manchola-Glez. 2017 (274)	72	0,80	[0,70 - 0,87]	7,20
Podnar 2017 (230)	46	0,75	[0,59 - 0,85]	6,57
Erdim 2019 (283)	71	0,91	[0,86 - 0,94]	7,18
Isa 2019 (236)	154	0,83	[0,77 - 0,87]	7,87
Zishan 2019 (284)	20	0,94	[0,85 - 0,98]	4,86
Pratanaphon 2020 (285)	115	0,67	[0,55 - 0,76]	7,66
Total (efectos aleatorios)	1464	0,82	[0,76 - 0,87]	100,00

N: tamaño de la muestra, ICC: índice de correlación Interclase; IC 95%: índice de confianza al 95%

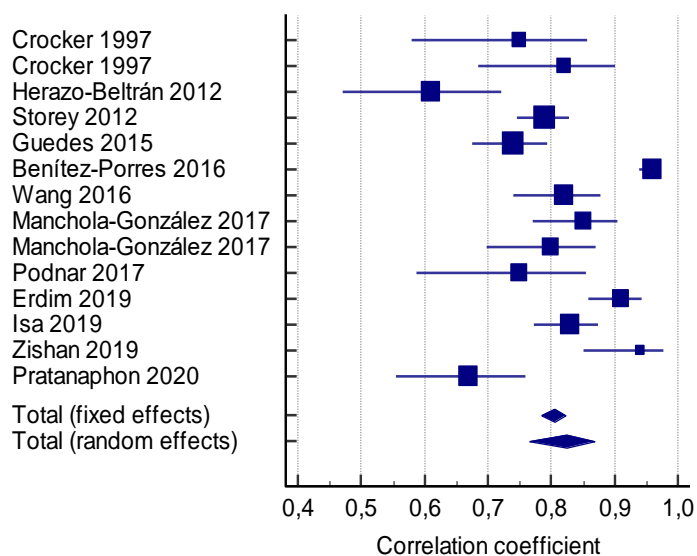


Figura 18. Forest-plot del meta-análisis de todos los estudios de fiabilidad del PAQ-C

Tabla 10. Meta-análisis de los estudios de fiabilidad del PAQ-C que han obtenido una calificación de la calidad metodológica de al menos "dudosa".

Estudio	N	ICC	IC 95%	Efectos aleatorios
Crocker 1997 (253)	43	0,75	[0,58 - 0,86]	8,43
Crocker 1997 (253)	41	0,82	[0,69 - 0,90]	8,26
Herazo-Beltrán 2012 (272)	100	0,61	[0,47 - 0,72]	10,87
Storey 2012 (232)	323	0,79	[0,75 - 0,83]	12,66

Guedes 2015 (278)	232	0,74	[0,68 - 0,79]	12,31
Wang 2016 (262)	92	0,82	[0,74 - 0,88]	10,68
Manchola-Glez. 2017 (274)	72	0,80	[0,70 - 0,87]	10,04
Erdim 2019 (283)	71	0,91	[0,86 - 0,94]	10,00
Zishan 2019 (284)	20	0,94	[0,85 - 0,98]	5,56
Pratanaphon 2020 (285)	115	0,67	[0,55 - 0,76]	11,18
Total (efectos aleatorios)	1109	0,79	[0,73 - 0,84]	100,00

N: tamaño de la muestra, ICC: índice de correlación Interclase; IC 95%: índice de confianza al 95%

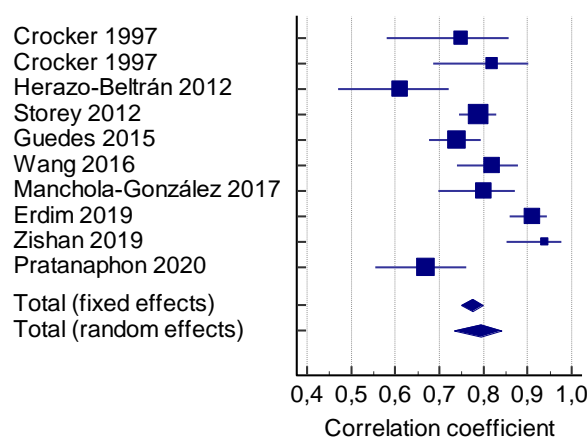


Figura 19. Forest-plot del meta-análisis de los estudios de fiabilidad del PAQ-C que han obtenido una calificación de la calidad metodológica de al menos "adecuada".

3.4.4. Error de medición

No se pudo evaluar el error de medición del EY-PAQ ya que se no se encontraron estudios que presentaran esta propiedad de medición.

Para el PAQ-C, se analizaron un total de 5 estudios correspondientes a 3 artículos que proporcionaban resultados del error de medición entre el test y el retest. En la Tabla 11 se muestran las características de estos estudios con respecto al intervalo entre el test y el retest, la calificación desglosada y general de la calidad metodológica del estudio, el tamaño de la muestra, el resultado y la calificación del resultado. Tal y como se ha adelantado en el

apartado de material y métodos, la calificación del resultado según los criterios para buenas propiedades de medición fue indeterminada (?) ya que no se conoce la diferencia mínima significativa.

Para el estudio de Guedes et al. (278) se calculó el error estándar de medición (SEM) a partir de los límites del acuerdo (LoA) presentados en los gráficos de Bland y Altman y utilizando la ecuación $SEM = SD \sqrt{1 - ICC}$, mientras que para obtener el SEM del estudio de Manchola-González et al. (274) se contactó directamente con los autores.

Tabla 11. Estudios de error de medición del PAQ-C Evaluación de la calidad metodológica y calificación del resultado

	Estudio	Intervalo test-retest	Calidad metodológica				N	Resultado	Calificación
			RD	AE	O	G			
PAQ-C	Guedes 2015 (278)	7 días	2	4	4	2	124	LoA: 0,88 y - 0,76 SEM = 0,23	(?) (?)
	Guedes 2015 (278)	7 días	2	4	4	2	108	LoA: 0,64 y - 0,70 SEM = 0,17	(?) (?)
	Manchola-Glez. 2017 (274)	2 horas	1	4	1	1	72	SEM = 0,15	(?)
	Manchola-Glez. 2017 (274)	7 días	2	4	4	2	72	SEM = 0,19	(?)
	Pratanaphon 2020 (285)	2 semanas	2	4	4	2	115	SEM = 0,36 SDC = 0,99	(?) (?)

PAQ-C: Physical Activity Questionnaire – Older Children, RD: requisitos de diseño, AE: análisis estadístico, O: otros, G: general, 4: “excelente”, 3: “adecuado”, 2: “dudoso”, 1: “inadecuado”, N: tamaño de muestra, LoA: límites del acuerdo, SEM: error estándar de medición, SDC: cambio más pequeño detectable

3.4.5. Validez de constructo

En la Tabla 12 se muestran las características de los estudios que evalúan la validez de constructo del EY-PAQ y el PAQ-C con respecto al instrumento

comparador, el constructo que mide, la calidad metodológica, el resultado y la calificación del resultado según los niveles de evidencia.

La validez de constructo para el EY-PAQ se evalúa en el estudio de validación original por comparación de métodos con acelerometría triaxial. Por un lado, se presentan los resultados para los ítems que preguntan sobre la AFMV y, por otro, los resultados para las preguntas sobre CS. En ambos casos, se muestra la correlación entre el acelerómetro y todos los datos del cuestionario y entre el acelerómetro y los datos que se encuentran dentro de los límites específicos de AFMV y CS definidos por Hnatiuk et al. (286). Esta es una estrategia que utilizan los autores para ajustar los errores de notificación en base al rango de estimaciones publicadas de CS y AFMV en el tiempo de vigilia en niños pequeños. Establecieron como límites el 30-94% del tiempo para el CS y el 2-41% para la AFMV (224). Al no haberse publicado ningún otro estudio sobre la validez de constructo del EY-PAQ, se determinó que el resultado era insuficiente y la calidad de la evidencia se calificó como muy baja.

Para el PAQ-C, no se consideró apropiado meta-analizar todos los estudios disponibles porque los instrumentos comparadores utilizados miden constructos diferentes. Sin embargo, se agruparon los estudios de validez de constructo disponibles por niveles de evidencia.

En primer lugar, se analizó el subgrupo de estudios que no aportaba evidencia de calidad suficiente según los criterios de Hidding et al. (13). En este caso, se compendiaron los estudios que comparaban el PAQ-C con métodos indirectos que miden un constructo diferente a la AFMV (3 estudios correspondientes a 1 artículo) (Tabla 13 y Figura 20). Se obtuvo un coeficiente de correlación total de 0,52 (IC 95% [0,38 – 0,62]) con un porcentaje de inconsistencia del 44,62% (IC 95% [0,00 – 83,52]). No se calificó el resultado compendiado ya que no se disponía de referencias acerca de la correlación esperada.

En segundo lugar, se presentó el resultado combinado de los estudios que ofrecen un nivel de evidencia 3, es decir, aquellos que utilizan métodos indirectos que miden el mismo constructo que el PAQ-C (3 estudios correspondientes a 2 artículos) (Tabla 14 y Figura 21). Se obtuvo un coeficiente de correlación total de 0,51 (IC 95% [0,43 - 0,58]) con un porcentaje de inconsistencia del 18,80% (IC 95% [0,00 - 97,78]). Los resultados se puntuaron como insuficientes (-) y la calidad de la evidencia se degradó en un nivel (calidad de la evidencia de nivel 3 moderada) por existencia de riesgo de sesgo grave.

En tercer lugar, se presentó la síntesis de los resultados que evalúan la validez de constructo por comparación con métodos directos que miden constructos diferentes a la AFMV, aunque estrechamente relacionados y que, en su conjunto, pueden aportar un nivel de evidencia 2 (5 estudios correspondientes a 5 artículos) (Tabla 15 y Figura 22). El resultado de este meta-análisis presentó un coeficiente de correlación de 0,33 (IC 95% [0,26 - 0,39]) con un porcentaje de inconsistencia del 0% (IC 95% [0,00 - 78,18]). Se decidió realizar dos nuevos subanálisis dentro de este subgrupo en el que se combinaron, por un lado, los estudios que analizaban la validez de constructo por comparación con métodos directos uniaxiales (2 estudios correspondientes a 2 artículos) y, por otro, los estudios que utilizaban instrumentos triaxiales (3 estudios correspondientes a 3 artículos) (Tabla 16 y Figura 23). Para el meta-análisis de los resultados de métodos uniaxiales, se obtuvo un coeficiente de correlación de 0,30 (IC 95% [0,21 - 0,38]) con un porcentaje de inconsistencia del 0% (IC 95% [0,00 - 0,00]). Para el meta-análisis que evaluaba la validez de constructo a través de instrumentos triaxiales, se obtuvo un coeficiente de correlación de 0,36 (IC 95% [0,26 - 0,45]) con un porcentaje de inconsistencia del 0% (IC 95% [0,00 - 96,48]) (Tabla 17 y Figura 24). Los resultados se puntuaron como insuficientes (-) en ambos casos y la calidad de la evidencia se degradó en un nivel (calidad de la evidencia de nivel 2 moderada) por existencia de riesgo de sesgo grave.

Por último, se realizó el meta-análisis de los estudios que aportan un nivel de evidencia 1 porque comparan el PAQ-C con los datos de métodos directos que miden el mismo constructo, es decir, la AFMV (5 estudios correspondientes a 5 artículos) (Tabla 18 y Figura 25). El resultado de este meta-análisis presentó un coeficiente de correlación de 0,34 (IC 95% [0,23 – 0,44]) con un porcentaje de inconsistencia del 53,45% (IC 95% [0,00 – 82,86]). Se estableció que existía inconsistencia moderada los criterios de Higgins y Thompson (268,269). Los resultados se puntuaron como insuficientes (-) y la calidad de la evidencia se degradó en un nivel (calidad de la evidencia de nivel 1 moderada) por inconsistencia.

Tabla 12. Estudios de validez de constructo. Evaluación de la calidad metodológica y calificación del resultado

	Estudio	Instrumento comparador	Constructo	Calidad metodológica				N	Resultado	Nivel de evidencia (calificación)
				RD	AE	O	G			
EY-PAQ	Bingham 2016 (224)	Acelerómetro triaxial	AFMV	4	4	4	4	149	r = 0,03	1 (-)
		Acelerómetro triaxial	AFMV (corregido)	4	4	4	4	102	r = 0,30	1 (-)
		Acelerómetro triaxial	CS	4	4	2	2	147	r = 0,02	3 (-)
		Acelerómetro triaxial	CS (corregido)	4	4	2	2	89	r = 0,19	3 (-)
PAQ - C	Kowalski 1997 (208)	Calificación de actividad (comparación con pares)	AF general	1	3	1	1	89	r = 0,63	--
	Kowalski 1997 (208)	Calificación de actividad (informe profesores)	AF general	1	3	1	1	88	r = 0,45	--
	Kowalski 1997 (208)	<i>7-day physical activity recall (7-DPAR)</i>	AF total	1	3	1	1	92	r = 0,46	--
	Kowalski 1997 (208)	Diario de AFMV	AFMV total	4	3	2	2	72	r = 0,51	3 (-)
	Kowalski 1997 (208)	<i>The Godin and Shepard Questionnaire</i>	AFMV total	1	3	2	1	97	r = 0,41	3 (-)
	Thomas 2014 (276)	<i>Self-Report Habit Index (SRHI)</i>	Hábito de AFMV	4	4	2	2	336	r = 0,55	3 (-)
	Kowalski 1997 (208)	Acelerómetro uniaxial	AF total	4	3	2	2	73	r = 0,39	2 (-)
	Storey 2012 (232)	Podómetro	AF total	4	4	1	1	342	r = 0,28	2 (-)
	Guedes 2015 (278)	Acelerómetro triaxial	AF total	4	4	2	2	232	r = 0,40	2 (+)
	Benítez-Porres 2016 (273)	Acelerómetro triaxial	pasos/día	4	4	2	2	78	r = 0,24	2 (-)
	Martín-Bello 2020 (287)	Acelerómetro triaxial	AF total	4	4	2	2	24	r = 0,30	2 (-)
	Guedes 2015 (278)	Acelerómetro triaxial	min AFMV	4	4	4	4	232	r = 0,48	1 (-)
	Benítez-Porres 2016 (273)	Acelerómetro triaxial	AFMV total	4	4	4	4	78	r = 0,28	1 (-)

Wang 2016 (262)	Acelerómetro triaxial	AFMV/día	4	4	4	4	385	r = 0,33	1 (-)
Gobbi 2016 (280)	Acelerómetro triaxial	AFMV total	4	4	4	4	55	r = 0,30	1 (-)
Ben Jemaa 2018 (282)	Acelerómetro triaxial	AFMV/día	4	3	4	3	40	r = 0,12	1 (-)

EY-PAQ: Early Years – Physical Activity Questionnaire, PAQ-C: Physical Activity Questionnaire – Older Children, RD: requisitos de diseño, AE: análisis estadístico, O: otros, G: general, r: correlación, AF: actividad física, AFMV: actividad física moderada-vigorosa, CS: comportamiento sedentario

Tabla 13. Meta-análisis de los estudios de validez de constructo que no aportan evidencia de suficiente calidad

Estudio	N	r	IC 95%	Efectos aleatorios
Kowalski 1997	89	0,63	[0,48 - 0,74]	33,19
Kowalski 1997	88	0,45	[0,26 - 0,60]	32,98
Kowalski 1997	92	0,46	[0,28 - 0,60]	33,83
Total (efectos aleatorios)	269	0,52	[0,38 - 0,62]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

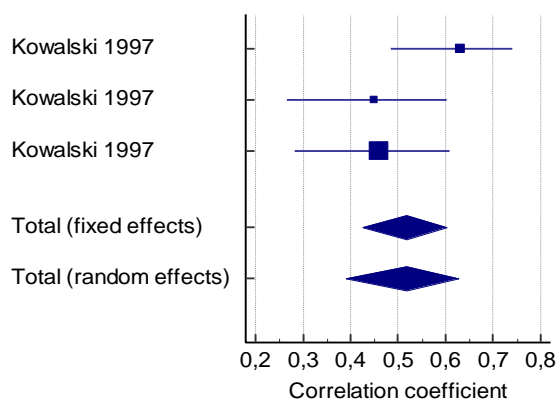


Figura 20. Forest-plot de los estudios de validez de constructo que no aportan evidencia de suficiente calidad

Tabla 14. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 3

Estudio	N	r	IC 95%	Efectos aleatorios
Kowalski 1997	72	0,51	[0,31 - 0,66]	17,68
Kowalski 1997	97	0,41	[0,22 - 0,56]	23,12
Thomas 2014	336	0,55	[0,47 - 0,62]	59,20
Total (efectos aleatorios)	505	0,51	[0,43 - 0,58]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

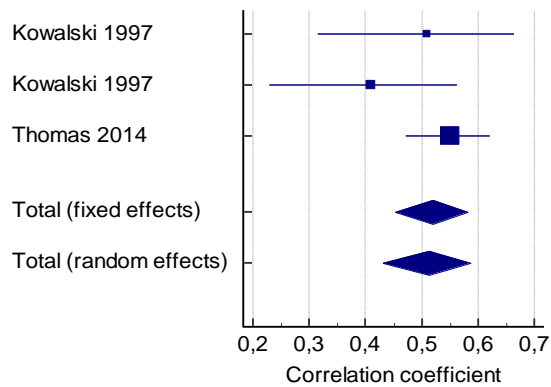


Figura 21. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 3

Tabla 15. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2

Estudio	N	r	IC 95%	Efectos aleatorios
Kowalski 1997	73	0,39	[0,17 - 0,56]	9,54
Storey 2012	342	0,28	[0,17 - 0,37]	46,19
Guedes 2015	232	0,40	[0,28 - 0,50]	31,20
Benítez-Porres 2016	78	0,24	[0,019 - 0,43]	10,22
Martín-Bello 2020	24	0,30	[-0,12 - 0,62]	2,86
Total (efectos aleatorios)	749	0,33	[0,26 - 0,39]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

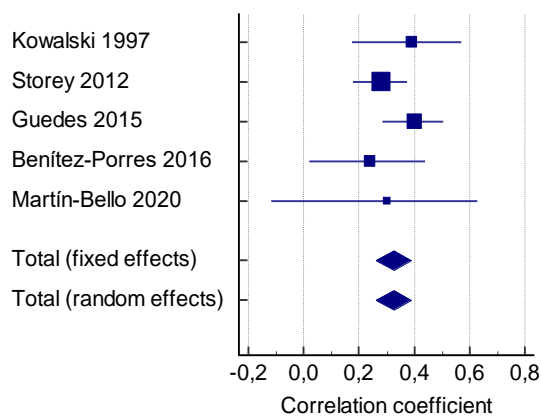


Figura 22. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2

Tabla 16. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos uniaxiales

Estudio	N	r	IC 95%	Efectos aleatorios
Kowalski 1997	73	0,39	[0,17 - 0,56]	17,11
Storey 2012	342	0,28	[0,17 - 0,37]	82,89
Total (efectos aleatorios)	415	0,30	[0,21 - 0,38]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

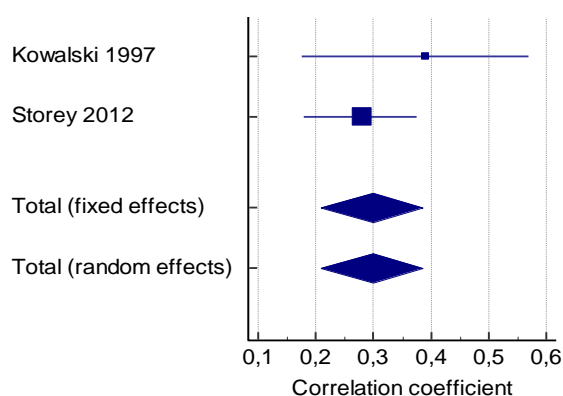


Figura 23. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos uniaxiales

Tabla 17. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos triaxiales

Estudio	N	r	IC 95%	Efectos aleatorios
Guedes 2015	232	0,40	[0,28 - 0,50]	70,46
Benítez-Porres 2016	78	0,24	[0,02 - 0,43]	23,08
Martín-Bello 2020	24	0,30	[-0,12 - 0,62]	6,46
Total (efectos aleatorios)	334	0,36	[0,26 - 0,45]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

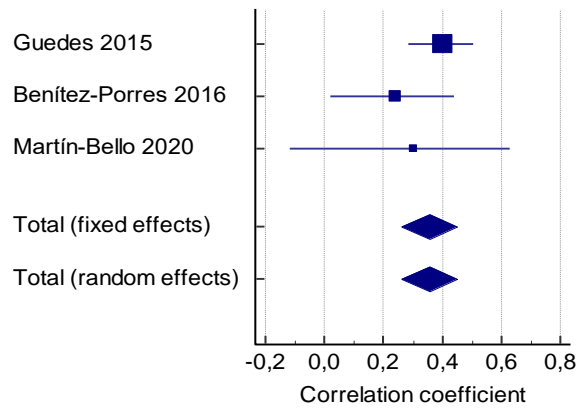


Figura 24. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 2 mediante comparación con métodos directos triaxiales

Tabla 18. Meta-análisis de los estudios de validez de constructo que aportan un nivel de evidencia 1

Estudio	N	r	IC 95%	Efectos aleatorios
Guedes 2015	232	0,48	[0,37 - 0,57]	27,81
Benítez-Porres 2016	78	0,28	[0,06 - 0,47]	16,68
Wang 2016	385	0,33	[0,23 - 0,41]	31,97
Gobbi 2016	55	0,30	[0,03 - 0,52]	13,20
Ben Jemaa 2018	40	0,12	[-0,19 - 0,41]	10,35
Total (efectos aleatorios)	790	0,34	[0,23 - 0,44]	100,00

N: tamaño de muestra, r: coeficiente de correlación, IC 95%: intervalo de confianza al 95%

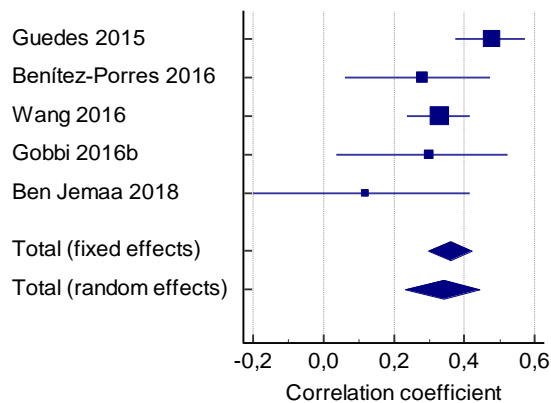


Figura 25. Forest-plot de los estudios de validez de constructo que aportan un nivel de evidencia 1

3.5. Discusión

Esta revisión sistemática resume la calidad de la evidencia para las propiedades de medición de los cuestionarios EY-PAQ para menores de 5 años y PAQ-C para escolares entre 8 y 14 años. Estos cuestionarios varían en sus atributos cualitativos y en la extensión del cuerpo de evidencia que ha estudiado sus propiedades de medición, siendo el PAQ-C un cuestionario ampliamente traducido y adaptado transculturalmente, mientras que para el EY-PAQ, por el momento, solo se dispone del estudio de validación original.

Las evaluaciones de la calidad metodológica, así como la puntuación según los criterios para buenas propiedades de medición mostraron que, en ninguno de los estudios incluidos en esta revisión, se obtuvo un resultado calificado como suficiente y además una calidad metodológica “excelente” o “adecuada”. Así mismo, el agrupamiento de los resultados puso en evidencia la gran heterogeneidad de los estudios en cuanto a sus características metodológicas y sus resultados, por lo que no fue posible concluir la suficiencia o insuficiencia de ninguna de las propiedades de medición con un nivel de evidencia alto.

Por otro lado, cabe destacar la dificultad para evaluar otras propiedades de medición, como la validez de contenido y la validez transcultural. En el primer caso, la falta de información sobre los procesos de desarrollo y las consultas formales a expertos y/o a muestras representativas de la población diana impidieron estudiar la adecuación de la validez de contenido según las recomendaciones de los expertos (247,250). Para la validez transcultural, el desafío es más complejo ya que los criterios planteados para evaluar esta propiedad de medición difieren del modelo formativo de los cuestionarios de AF, por lo que su calificación se puede basar únicamente en un juicio cualitativo en base a la información proporcionada por los autores sobre los procesos de traducción, síntesis, retrotraducción, revisión y comprobación en la población diana (201).

Por último, cabe destacar la ausencia de estudios que exploren la capacidad de respuesta en ninguno de los cuestionarios incluidos en esta revisión.

3.5.1. Validez de contenido

La forma más extendida de evaluar la validez de contenido es a través de entrevistas con profesionales y muestras representativas de la población diana para indagar sobre la relevancia, exhaustividad y comprensibilidad de cada uno de los ítems (243,244). En este contexto, el estudio de validación original del EY-PAQ no incluyó ningún tipo de consulta formal con expertos ni muestras representativas de la población (224). Para el PAQ-C, se menciona que la versión preliminar del cuestionario se modificó en base a las aportaciones del equipo investigador y una pequeña muestra de estudiantes entre 9 y 14 años, pero no se concreta el procedimiento de las consultas (253). Por tanto, no fue posible evaluar esta propiedad de medición en ninguno de los cuestionarios incluidos (247,250).

En general, la validez de contenido de los cuestionarios de AF está poco desarrollada. Otras revisiones sistemáticas centradas en cuestionarios de AF para la franja de pediatría mostraron las mismas dificultades en relación a la evaluación de esta propiedad de medición (9,13).

3.5.2. Validez transcultural

A diferencia de los métodos directos de medición de la AF, los cuestionarios son culturalmente dependientes, es decir, los resultados de validez evaluados en una población no pueden extrapolarse sistemáticamente a otras poblaciones, grupos étnicos o regiones geográficas, sino que se requiere de un proceso de adaptación del instrumento que asegure su aplicabilidad en cada contexto (201,249).

Este proceso consta de diferentes fases entre las que se incluye la traducción, síntesis, retrotraducción, revisión y comprobación en una muestra representativa de la población diana (201,249). Aunque estrictamente el proceso de traducción no es una propiedad de medición, condiciona el éxito de la validez transcultural en el idioma de destino, por lo que ha de cumplir unos mínimos requisitos para garantizar su calidad. De hecho, algunos expertos señalan que la baja validez transcultural de algunos instrumentos está justificada por la traducción deficiente, la mala adecuación de los ítems a la cultura de destino y la baja especificidad de las actividades preguntadas en el cuestionario para la nueva población diana (249). Además, se ha demostrado que cuando una comunidad no interpreta las preguntas en el sentido en el que fueron formuladas, se introduce un nuevo tipo de error en la medida que puede sesgar las diferencias verdaderas en los comportamientos de AF (288).

Los expertos en validación de medidas indirectas proponen que el proceso de traducción se inicie con la colaboración de al menos dos traductores bilingües independientes que traduzcan el cuestionario del idioma original al idioma de destino (201,289). La mayoría de los estudios incluidos en esta revisión mencionan que participaron dos expertos en el proceso de traducción (275,277,278,280,284,285,290), mientras que en otros se involucraron un total de 3 traductores (262,283), 1 traductor (236) y en algunos no se concreta el número exacto (271,279). En este sentido, cabe destacar que la traducción por parte de un único experto es el enfoque más sencillo aunque está limitado por la dependencia de las habilidades lingüísticas y de interpretación de un solo individuo, lo que a menudo conduce a una baja validez de la medida traducida (249). Por último, en algunos artículos se sobreentiende que se realizó un procedimiento de traducción aunque no se proporciona evidencia clara (281,282).

El siguiente paso consiste en combinar los resultados y sintetizar las versiones traducidas, preferiblemente con ayuda de un tercer investigador (201,289). En la mayoría de los artículos incluidos, se sobreentiende que se compendiaron las propuestas de los traductores aunque solo se menciona de manera específica para las versiones del PAQ-C al croata (275), al portugués de Brasil (278), al gujaratí (284) y al italiano (285). Con respecto al número de miembros involucrados en la síntesis de las versiones traducidas, varía de los dos traductores iniciales (275), a la colaboración de tres (278,285) e incluso cinco expertos (284).

A continuación, la versión traducida debe someterse a un proceso de traducción inversa por parte de nuevos traductores bilingües cegados a la versión original que puedan detectar errores derivados de la traducción (201,289). En este sentido, los estudios mencionados cuentan con la colaboración de uno (236,280) dos (273,277-279,283,285) o cinco traductores independientes (284), mientras que en algunos estudios no se aclara el número de retrotraductores implicados (262,271).

Una vez finalizada la traducción inversa, el comité de expertos que participan en el estudio de validez transcultural debe revisar todas las traducciones, tomar decisiones sobre las discrepancias generadas y redactar la versión preliminar del cuestionario (201,289). Cabe destacar que la simple traducción de un instrumento de un idioma a otro no conduce a una equivalencia lingüística y conceptual (291) por lo que es fundamental que el equipo involucrado en la adaptación del cuestionario conozca en detalle la cultura de la comunidad de destino (249). En los estudios de validez transcultural del PAQ-C, solo en algunos casos se menciona la composición de un comité de expertos formado por tres (271), cinco (277), siete (285), nueve (278) o trece expertos (283).

Por último, el cuestionario debe someterse a una prueba en una muestra representativa de la población diana cuyo objetivo es corroborar la

comprensibilidad, relevancia y exhaustividad de cada uno de los ítems propuestos (201,289). En relación a los estudios de validez transcultural del PAQ-C, solo en algunos casos se menciona que se haya realizado algún tipo de consulta formal a la población diana. Por ejemplo, en la adaptación del cuestionario al persa se llevó a cabo una entrevista cognitiva en 4 estudiantes (271), mientras que en el estudio de validación en Tailandia se realizaron grupos focales de 3-5 participantes (285). Sin embargo, el método más habitual fue la realización de pruebas piloto en muestras representativas que variaron de 10 (236,276), 47 (278), 50 (272) o 84 niños (283).

Cabe destacar que las modificaciones del PAQ-C con respecto a la versión original publicada por Crocker y Kowalski (208,253) afectan principalmente al listado de actividades incluidas en el ítem 1. Sin embargo, algunas versiones incorporan otras modificaciones además de la revisión de las actividades del checklist inicial. Por ejemplo, Moore et al. adaptaron el PAQ-C para una población de niños afroamericanos y europeos americanos de entre 8 y 9 años. Como la muestra era de menor edad en comparación con estudios anteriores decidieron omitir el ítem 4 (270). Por otro lado, en la versión colombiana del PAQ-C se observó que para los niños de 8 años era confusa la pregunta sobre la AF durante el almuerzo, por lo que se aclaró que eran las actividades realizadas antes y después de comer (272). Así mismo, para crear la versión tailandesa se incluyeron cambios menores en los datos de identificación del encuestado y aclaraciones en el enunciado de las preguntas (por ejemplo *"por favor complete lo siguiente como lo hace habitualmente"*) o las opciones de respuesta (por ejemplo, en la pregunta 2 se añadió un porcentaje a cada adverbio de frecuencia: *nada (0%), casi nunca (20%), a veces (<50%), con bastante frecuencia (> 50%) y siempre (80%)*) (285). No se incluyó en esta revisión el estudio de Janz et al. del 2008 ya que se consideró que las modificaciones que hacen los autores para adaptar el PAQ-C a los períodos no lectivos son lo suficientemente sustanciales como para

considerarlo un nuevo cuestionario que requiere un proceso de desarrollo y validez de contenido independiente (292).

Una vez finalizado el análisis cualitativo de los procesos de traducción, se inicia el estudio de validez transcultural como parte del proceso de comprobación de que la nueva versión del cuestionario sea un reflejo adecuado de la representación de los ítems de la versión original (239). Por tanto, es evidente que se requiere una descripción detallada de las características relevantes y la puntuación obtenida por al menos dos subgrupos diferentes, en este caso, la muestra de participantes en el estudio de validación original y la muestra de población cultural o idiomáticamente diferente a la que se quiere adaptar el cuestionario (247). Para los cuestionarios construidos bajo un modelo reflexivo existen métodos estadísticos ampliamente reconocidos que evalúan si los ítems son invariantes en la medición, como el Análisis Factorial Confirmatorio (AFC), el análisis de regresión logística o el Funcionamiento Diferencial de Ítems (DIF) (201,250). Todos estos análisis pretenden comprobar si los encuestados de distintos grupos culturales, pero con los mismos rasgos latentes, responden de manera similar a un ítem particular o al cuestionario en general. Por ello, es imprescindible que, excepto por la variable de grupo, ambas muestras sean similares en las características relevantes (247). En el caso de los cuestionarios de AF, los métodos estadísticos mencionados son irrelevantes y todavía no se han planteado teorías de medición que expliquen las relaciones entre el constructo y los ítems. Por tanto, el análisis de las similitudes entre los resultados de la validación original y los resultados de las versiones adaptadas solo puede hacerse desde un punto de vista cualitativo en el que se compare la edad, la distribución por sexos y la puntuación del cuestionario para cada muestra de participantes (228).

En este contexto, los estudios de traducción y adaptación transcultural del PAQ-C deben tomar como referencia los artículos de validación original

publicados en Canadá en 1997 por Crocker y Kowalski (208,253). Estos artículos presentan un total de 3 estudios realizados en muestras de niños entre 8 y 15 años con una media de edad que ronda los 11 años y una distribución homogénea por sexos. La puntuación del PAQ-C se sitúa alrededor de los 3 puntos, siendo ligeramente superior para los niños en comparación con las niñas (253).

A partir de estas referencias, los estudios de validez transcultural varían en su diseño metodológico en cuanto a las características relevantes de la muestra. Por un lado, los artículos de validación al portugués de Brasil (278), al español (273,274), al chino (262), al turco (283), al gujaratí (284) y al tailandés (285) fueron realizados en muestras de población de un rango de edad muy similar a los estudios de validación original, reclutando participantes entre 8-13 (262,273,278), 8-14 (274,284,285) y 9-14 años (283), que abarcan la amplitud de la franja de edad a la que se dirige el PAQ-C. La media de edad de estos estudios es similar a la muestra de Crocker y Kowalski (208,253), excepto para el estudio de Zishan et al. (284) en el que la media asciende a 12,75 (1,29) años. Por otro lado, algunos estudios utilizan muestras más acotadas que no representan la totalidad de la población diana. Por ejemplo, el artículo de adaptación a EE.UU. utiliza una muestra de 8-9 años (270), los artículos de validación al italiano y al árabe administran el cuestionario a niños entre 8-11 años (280,282), los estudios de adaptación al español de Colombia y al inglés de Reino Unido emplean muestras de participantes de 9-11 años (272,276), el estudio de validación al japonés utiliza un rango de edad de entre 9-12 años (236) y el estudio de validación al croata reduce su muestra a participantes de 10 años (275). Por último, cabe destacar que algunos de los artículos incluidos en la síntesis cualitativa para la validez transcultural fueron excluidos de los análisis cuantitativos posteriores por utilizar muestras de participantes que se alejan de población de interés para el PAQ-C, por lo que se cuestiona su validez en la representación de los datos. En este grupo se incluyen los estudios de

validación al persa (271), al holandés (277), al malayo (279) y al afrikáans (281).

Con respecto a la distribución por sexos, casi todos los estudios presentan muestras más o menos equitativas que no sobrepasan la relación 4:6, a excepción de los estudios de validación al persa (271) y al malayo (279), que presentan un mayor porcentaje de niñas, siendo del 75,61% y del 65,75%, respectivamente.

En relación al tamaño de la muestra, la mayor parte de las recomendaciones publicadas hacen referencia a los cálculos propios de modelos reflexivos, por lo que no son aplicables en este contexto (247). Sin embargo, existen pautas generales, como las establecidas por Sousa et al. en el 2011, quienes sugieren administrar el cuestionario a un mínimo de 10 individuos por cada ítem (293). Siguiendo la propuesta de estos autores, el tamaño de muestra mínimo recomendado para los estudios de validez transcultural del PAQ-C es de 90 individuos, requisito que cumplen solo algunos de los estudios incluidos (236,262,270,272,276-278,280,283,285). Además, la guía COSMIN aclara que es una condición importante ya que los resultados sobre esta propiedad de medición son únicos y no pueden agruparse (247,250). Como excepción, el estudio de validez transcultural del PAQ-C a la población española fue llevado a cabo por dos grupos de investigación diferentes (273,274). Se consultó a los autores el motivo de esta discrepancia y explicaron que en el momento de iniciar el proyecto no había ninguna validación al español publicada y que paralelamente ambos grupos trabajaron con el mismo objetivo, siendo el estudio de Benítez-Porres et al. aceptado para publicación en marzo de 2016 y publicado en septiembre de 2016 (273), mientras que el estudio de Manchola et al. fue aceptado para publicación en noviembre de 2015 y finalmente publicado en la revista en marzo de 2017 (274).

En relación a la puntuación del PAQ-C, la mayor parte de los estudios presentan resultados globales muy similares a los informados en la validación

original (208,253). Así mismo, los estudios que reportan los datos diferenciados por sexos coinciden en mostrar mayores resultados para los niños, a excepción de los estudios realizados en Brasil y en España, que presentan puntuaciones ligeramente superiores para las niñas (273,278). Los estudios cuyas puntuaciones totales difieren con respecto a las reportadas por Crocker y Kowalski (208,253) coinciden en presentar resultados indicativos de un menor nivel de AF. El ejemplo más claro es el estudio de Herazo-Beltrán et al., que muestra una calificación global de 1,90 (0,3) puntos, lo que supone casi un punto y medio de diferencia con los estudios originales (272). Así mismo, destacan las bajas puntuaciones obtenidas en las muestras de Brasil (278), China (262), España (274), Túnez (282), Japón (236) y Tailandia (285), que se sitúan entre los 2 y 3 puntos. Sin embargo, no se han encontrado claras diferencias ni en la distribución por sexos ni en la edad de la muestra que puedan justificar las bajas puntuaciones obtenidas en estos ejemplos. Sin datos adicionales que aporten más información acerca de las características socioculturales de estas poblaciones es difícil discernir si el motivo de las bajas puntuaciones se justifica por las diferencias en la realidad cultural de estos países o es debido un sesgo en el reclutamiento de la muestra (201).

3.5.3. Fiabilidad

La fiabilidad se refiere al grado en que las puntuaciones de los respondedores son las mismas en mediciones repetidas siempre y cuando se hayan mantenido las condiciones de medición (239). Los requisitos sobre su planteamiento deben ser estrictos en lo que se refiere a la independencia de las administraciones y la estabilidad tanto del contexto como de los participantes. Por ello, es fundamental encontrar el equilibrio en el intervalo entre el test y el retest para evitar que los sujetos recuerden las respuestas

ofrecidas en primera instancia, pero asegurando el mantenimiento del contexto de medición y el estado del participante (177,193,247).

Con respecto a la calidad metodológica de los estudios que evalúan la fiabilidad del EY-PAQ y el PAQ-C, en ninguno de los casos se proporcionaba clara evidencia de que los participantes se mantuvieran estables entre la primera y segunda administración y solo en los artículos de Guedes et al. (278) e Isa et al. (236) se menciona que las condiciones del test fueron iguales en ambas ocasiones. Por ello, todos los estudios vieron rebajada su calidad metodológica al menos dos puntos. Adicionalmente, algunos estudios utilizaban intervalos test-retest que incumplían los criterios definidos para el período entre administraciones (236,273,274). En estos casos, la calidad metodológica se puntuó como “inadecuada”.

En relación a los métodos estadísticos, solo cuatro estudios aclaraban el modelo o fórmula utilizada para calcular el ICC y recibieron una calificación de “excelente” para este estándar (224,236,262,273). El resto de los estudios simplemente presentaban los resultados del ICC, por lo que fueron puntuados como “adecuados”. Así mismo, algunos estudios reportaban estadísticos de correlación a través del coeficiente de Pearson (284) o de Spearman (230,284). Sin embargo, en todos los casos se acompañaban también del ICC por lo que estos resultados se obviaron para homogeneizar los análisis.

En las revisiones sistemáticas de propiedades de medición de cuestionarios de AF para pediatría publicadas en 2010 y 2018, se discute la misma problemática en relación a la pobre metodología de los estudios (9,13). Aunque los autores de estas revisiones utilizan la versión del 2010 del Checklist de Riesgo de Sesgo de COSMIN (294), las limitaciones que expresan son similares. Sin embargo, es importante destacar que en su caso se penaliza la calidad metodológica de los estudios por cuestiones que ya no se contemplan en el checklist del 2018 (250). Por ejemplo, se rebaja la calidad

metodológica si no se proporciona información acerca del manejo de los datos faltantes o si el tamaño de muestra no es apropiado (294).

Con respecto a los resultados obtenidos en los estudios individuales, para el EY-PAQ los índices de concordancia muestran una pobre correlación entre las administraciones (224). En otros cuestionarios de AF dirigidos a preescolares, se observan resultados de fiabilidad muy variados. Por ejemplo, el *Preschool-age Children's Physical Activity Questionnaire* (Pre-PAQ) presenta un ICC bajo que varía de 0,44 a 0,64 a excepción de las actividades organizadas que muestran un ICC de entre 0,96 y 0,99 (295) o el *Kid Active Q*, con índices de fiabilidad entre 0,60 y 0,66 (296). Por otro lado, el CLASS para preescolares muestra resultados de fiabilidad muy diversos que van del 0,49 al 0,89 en función de si se analiza la duración o la frecuencia para las actividades de intensidad moderada, vigorosa o total (205). Por el contrario, el ERBs cumple con los criterios para buenas propiedades de medición al presentar un ICC de 0,71 - 0,93 para la dedicación deportiva y de 0,82 - 0,91 para el transporte activo (261).

Sin embargo, para el PAQ-C, la mayoría de los resultados son indicativos de una fiabilidad que supera los criterios establecidos en esta revisión, a excepción de dos de los estudios incluidos (272,285). En otros cuestionarios de AF dirigidos a edad escolar, los resultados de fiabilidad son muy variados. Por ejemplo, el cuestionario de AF del *South American Youth/Child Cardiovascular and Environment Study* (SAYCARE) muestra una fiabilidad insuficiente expresada a través de coeficientes de correlación de Spearman entre el test y el retest que varían de 0,28 a 0,89, siendo suficiente únicamente para las actividades de intensidad vigorosa (297). Por otro lado, tanto el *Godin Leisure-Time Exercise* original como su versión modificada presentan índices negativos de fiabilidad (260,298). En el primer caso, el coeficiente entre las administraciones es de 0,25 - 0,69, ascendiendo según se incrementa la intensidad de las actividades preguntadas (260). En el segundo caso, se muestra un único coeficiente de correlación equivalente a 0,68 (298).

Así mismo, la versión china del CLASS para escolares muestra resultados indeterminados, con índices de concordancia que varía del 0,61 al 0,71 según si se estudian los resultados de AF de intensidad moderada, vigorosa o total (299). Por el contrario, el cuestionario *Energy-child* presenta resultados suficientes de fiabilidad que se sitúan entre 0,70 y 1,00 para el transporte, la AF durante el recreo y las horas de deporte (237), al igual que el cuestionario *Girls health Enrichment Multisite Study Activity*, que muestra índices globales de suficiente fiabilidad (ICC = 0,70 -0,82) (255).

La tendencia a una mejor fiabilidad de los cuestionarios dirigidos a escolares en comparación con aquellos dirigidos a preescolares podría ser debida a varios factores. Por ejemplo, la influencia del respondedor en la percepción de la AF (7,172,300,301) o la mayor variabilidad de los patrones de movimiento en niños menores de 5 años (15,29,56,96).

En relación a la agrupación de los estudios del PAQ-C, aunque en la mayor parte de los casos la calificación de los resultados según los criterios para buenas propiedades de medición era positiva, la síntesis de todos ellos dio como resultado una alta inconsistencia, no resuelta a través de análisis por subgrupos. Por tanto, el resumen de la fiabilidad del PAQ-C fue calificado como suficiente, aunque con un nivel de evidencia muy bajo. Por un lado, de acuerdo a la perspectiva de otros autores, se consideró que, debido a la baja calidad metodológica de los estudios individuales, la calidad global debía degradarse de alta a moderada. Por otro, la alta inconsistencia del meta-análisis sugirió que la calidad debía rebajarse dos niveles más, es decir, de moderada a muy baja.

No se ha encontrado ninguna otra revisión sistemática que agrupe los resultados de fiabilidad publicados sobre un mismo cuestionario de AF, por lo que no podemos comparar nuestros resultados con los reportados por otros autores. Para el EY-PAQ, al no haberse publicado ningún estudio

adicional sobre sus propiedades de medición, se concluyó que la fiabilidad era insuficiente y la calidad de la evidencia se calificó como muy baja.

3.5.4. Error de medición

El error de medición representa la idea de que todas las observaciones están sometidas a variaciones derivadas de diferentes fuentes, como los instrumentos, los evaluadores, los sujetos sometidos a la medida o las circunstancias bajo las que se realizan las valoraciones (201,222,247). La estimación del error de medición está íntimamente relacionada con la evaluación de la fiabilidad test-retest de un cuestionario. Habitualmente, el protocolo para estudio del error de medición cumple las mismas pautas que el estudio de fiabilidad, por lo que se pueden extrapolar las condiciones de medición de un estudio a otro (200,247). Por ello, la calificación global de la calidad metodológica de los estudios de error de medición incluidos en esta revisión es igual a la asignada a los estudios de fiabilidad. Sin embargo, cabe destacar que el estándar en el Checklist de Riesgo de Sesgo de COSMIN que evalúa los métodos estadísticos para estimar el error de medición, asigna una puntuación de “excelente” si se presentan los resultados del SEM, el SDC o el LoA, mientras que este mismo estándar para los estudios de fiabilidad es más estricto ya que requiere que los autores aporten información sobre el modelo o fórmula utilizados (250). Debido a esta diferencia, la puntuación de la calidad metodológica de los métodos estadísticos para el error de medición obtuvo el máximo valor mientras que, para los estudios de fiabilidad, en la mayoría de los casos se asignó una calificación de “adecuado”.

En la revisión sistemática de propiedades de medición publicada en 2010 no se incluyó ningún estudio que evaluase el error de medición (9). Sin embargo, en su actualización más reciente se recoge un total de nueve estudios que evalúan esta propiedad (13). Al igual que ocurría con la calificación de la calidad metodológica de los estudios de fiabilidad, Hidding et al. (13)

manifiestan limitaciones similares a las expresadas en esta revisión aunque en su caso la aplicación de la versión del 2010 del Checklist de Riesgo de Sesgo de COSMIN es más restrictiva en cuanto a la degradación de la calidad (294).

Con respecto a los resultados obtenidos en los estudios individuales, para el EY-PAQ no se conocen datos sobre de error de medición, mientras que para el PAQ-C se analizaron 5 estudios correspondientes a 3 artículos (274,278,285). Sin embargo, no se pudo aportar una puntuación del resultado según los criterios para buenas propiedades de medición ni una calificación global de la evidencia porque se desconoce la diferencia mínima significativa.

3.5.5. Validez de constructo

La validez de constructo aporta evidencia en aquellas situaciones en las que no es posible comparar un instrumento indirecto con su *gold standard*. Aunque los resultados son menos concluyentes, si las hipótesis está bien fundamentadas pueden ofrecer un nivel de evidencia suficiente que permita determinar si el cuestionario proporciona las puntuaciones esperadas en base al constructo que pretende medir (201).

Por este motivo y de acuerdo con las recomendaciones de la guía COSMIN (247), se definieron hipótesis a priori acerca de las correlaciones esperadas entre los cuestionarios y los instrumentos comparadores. Esta es una de las principales diferencias entre el Checklist de Riesgo de Sesgo de COSMIN original (294) y el actualizado en el 2018 (250). En la primera versión, uno de los estándares valoraba si los autores de los estudios incluidos habían definido hipótesis acerca de los resultados esperados, lo que a menudo conducía a una calificación “inadecuada” (294). Con la actualización del checklist, al recomendar al equipo de revisión que formule las hipótesis, la calidad metodológica no depende del planteamiento de los autores y todos los resultados se pueden comparar bajo los mismos criterios (250). Así

mismo, al igual que ocurría con la calificación de la calidad metodológica de los estudios de fiabilidad y error de medición, el checklist del 2010 penalizaba la calificación de los estudios sobre validez de constructo por no proporcionar información acerca del manejo de los datos faltantes o realizar el estudio en una muestra insuficiente de participantes (294). Sin embargo, y a pesar de las diferencias en los estándares utilizados, los resultados de nuestra revisión y de las revisiones publicadas anteriormente coinciden en mostrar una pobre calidad metodológica para aquellos estudios que utilizan medidas de comparación con propiedades de medición desconocidas o insuficientes o cuyo constructo no se corresponde con el que pretende medir el cuestionario (9,13).

En relación a la calidad metodológica asignada al estudio de validez de constructo para los ítems sobre AFMV del EY-PAQ (224), se consideró calificarlo con la máxima puntuación en base a los criterios establecidos para el nivel de evidencia 1. Sin embargo, para las preguntas que evalúan el CS, la evidencia disponible se categorizó como nivel 3 ya que los datos aportados por los autores utilizan los puntos de corte de Costa et al., quienes establecen un límite de 5 recuentos cada 5 segundos, es decir, 60 cpm para definir el CS (302).

Por otro lado, los resultados de este estudio muestran correlaciones muy bajas entre el EY-PAQ y los resultados de un acelerómetro, siendo significativas únicamente para la AFMV corregida por límites ($r = 0,30$) (224). En otros cuestionarios de AF dirigidos a preescolares, se observan resultados similares al reportado por Bingham et al. para el EY-PAQ (224). Por ejemplo, se ha estudiado la validez de constructo por comparación de las puntuaciones de los cuestionarios con los recuentos de AFMV de acelerómetros triaxiales, dando como resultado correlaciones de 0,17 para el Pre-PAQ (295), 0,34 para el cuestionario de Burdette modificado (303), 0,10 para el cuestionario de Harro modificado (303) y 0,42 para el CPAQ (219).

Con respecto a los estudios de validez de constructo del PAQ-C destaca la gran variedad de diseños e instrumentos comparadores utilizados. Por ello, la calificación del estándar 4 (“¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?”) del checklist de Riesgo de Sesgo de COSMIN se definió como “excelente” si se estudiaba la correlación entre la puntuación del PAQ-C y los datos de AFMV de un acelerómetro triaxial (nivel de evidencia 1). Se consideró que existían fallos metodológicos menores si la validez de constructo se evaluaba por comparación contra recuentos de AF total de un acelerómetro o podómetro (nivel de evidencia 2) o contra métodos indirectos que miden el mismo constructo, es decir, AFMV (nivel de evidencia 3). Para los demás planteamientos se determinó que existían fallos importantes en el diseño del estudio.

En primer lugar, se presentó el meta-análisis de los resultados de validez de constructo por comparación del PAQ-C con instrumentos indirectos que miden constructos diferentes a la AFMV. Kowalski et al. (208) utilizan un índice de actividad por comparación con pares (304), un índice de actividad mediante el informe de profesores (305) y el *cuestionario 7- Day Physical Activity Recall (7-DPAR)* (305) que mide la AF en todos los rangos de intensidad. La degradación de la calidad metodológica en estos estudios se justificó por varios motivos. En primer lugar, se consideraron inadecuadas las propiedades de medición de todos los instrumentos comparadores seleccionados (304–308). En segundo lugar, la validez de constructo pierde valor en tanto que la puntuación del PAQ-C se compara con los resultados de instrumentos indirectos que miden constructos similares, pero no idénticos, por lo que se consideró que existían otros fallos metodológicos importantes (247).

El meta-análisis de este subgrupo presentó un resultado de correlación de 0,52 (IC 95% [0,38 – 0,62]) con un porcentaje de inconsistencia bajo entre los estudios incluidos. Según los criterios de Hidding et al. estos resultados no aportan un nivel de evidencia suficiente para sustentar la validez de

constructo (13), por lo que no se calificó el resultado agrupado y no se consideró realizar nuevos subanálisis.

Los estudios que correlacionan la puntuación del PAQ-C con otros métodos indirectos que miden el mismo constructo pueden proporcionar un nivel de evidencia 3. En este meta-análisis se incluyeron los estudios de Kowalski et al. (208) y Thomas et al. (276). Concretamente, utilizaron como instrumentos comparadores un diario que registra un total de 10 AFMV (309), el cuestionario *The Godin and Shepard Leisure Time Exercise*, que anota la frecuencia en la participación de AFMV (310) y el cuestionario *Self-Report Habit Index* (SRHI) que mide el hábito de AFMV (311). En base a estudios previamente publicados, se consideró que las propiedades de medición eran suficientes para el diario de actividad (312) y el SRHI (311,313) mientras que para el cuestionario *The Godin and Shepard Leisure Time Exercise* se calificaron como insuficientes (260). Con respecto al estándar que valora la existencia de otros fallos metodológicos, la calidad se rebajó a “dudosa” en base a los criterios definidos. Finalmente, para los estudios de validez de constructo de Kowalski et al. (208) se asumió que el procedimiento estadístico era apropiado aunque los autores no informaban del análisis realizado, por lo que el estándar 3 del checklist de COSMIN (250) se degradó de “excelente” a “adecuado” (250).

El resumen de los estudios de esta categoría dio como resultado una correlación total de 0,51 (IC 95% [0,43 - 0,58]) con un porcentaje de inconsistencia del 18% (IC 95% [0,00 - 97,78]), indicativo de homogeneidad entre los estudios. Sin embargo, cabe destacar que el IC 95% para la inconsistencia es muy amplio, por lo que se deduce que el cálculo asume un alto error, probablemente debido al bajo número de estudios incluidos. Por ello, la calidad de la evidencia se calificó como insuficiente y se rebajó en un nivel (de alta a moderada) debido a la existencia de riesgo de sesgo grave, pero no se consideró apropiado degradarla por inconsistencia a pesar de la gran amplitud del IC 95%.

Por otro lado, dos estudios presentaron correlaciones entre la puntuación del PAQ-C y los resultados de métodos directos uniaxiales. En primer lugar, el artículo de Kowalski et al. (208) utilizó como medida de comparación los recuentos de AF total medidos por un acelerómetro uniaxial (308). En segundo lugar, Storey et al. (232) estudiaron las asociaciones entre el PAQ-C y los recuentos de podómetros. Con respecto a las propiedades de medición de estos instrumentos, se consideró que eran suficientes en ambos casos (232,308,314). En relación al procedimiento estadístico, para el estudio de Kowalski et al. (208) no se presentaban los análisis realizados, por lo que se penalizó la calificación en un nivel (250). Con respecto a la existencia de otros fallos metodológicos, el estudio de Kowalski et al. (208) se puntuó como “dudoso”, en base a los criterios definidos, mientras que para el estudio de Storey et al. (232) la calidad se rebajó a “inadecuada” al considerar que el tiempo de medición del podómetro no aseguraba la representación completa del período de recuerdo del PAQ-C. Según la clasificación de Hidding et al., los resultados que presentan correlaciones entre la puntuación del PAQ-C y los métodos directos uniaxiales pueden aportar un nivel de evidencia 2 (13). Se obtuvo una correlación agrupada de 0,30 (IC 95% [0,21 - 0,38]) con homogeneidad entre los estudios y se calificó como insuficiente con calidad de la evidencia de nivel 2 moderada debido al riesgo de sesgo grave de los estudios incluidos.

Por último, se analizó la calidad metodológica de los estudios que valoran la validez de constructo por comparación con acelerometría triaxial. En todos los casos, se consideró que las propiedades de medición eran suficientes (315–320). Con respecto a los estudios que relacionan las puntuaciones del PAQ-C con los recuentos de AF total (273,278,287), se rebajó la calidad de la evidencia a “dudosa” en base a los criterios definidos, mientras que para los estudios que comparan la puntuación del PAQ-C con los recuentos de AFMV (262,273,278,280) se mantuvo la máxima puntuación. Como excepción, el estudio de Ben-Jemaa (282) vio reducida su calificación de “excelente” a

“adecuada” debido a que no se detalla el análisis estadístico realizado para estudiar la correlación entre el PAQ-C y el acelerómetro.

En el primer caso, se obtuvo un coeficiente de correlación de 0,36 (IC 95% [0,26 - 0,45]) con un porcentaje de inconsistencia del 0% (IC 95% [0,00 - 96,48]). Los resultados se puntuaron como insuficientes y la calidad de la evidencia se degradó en un nivel por existencia de riesgo de sesgo grave. En el segundo caso, se obtuvo una correlación de 0,34 (IC 95% [0,23 - 0,44]) con una inconsistencia del 53,45% (IC 95% [0,00 - 82,86]). Los resultados se consideraron insuficientes en base a los criterios definidos para buenas propiedades de medición y la calidad de la evidencia se calificó como nivel 1 moderada. En ambos subgrupos, los resultados sobre la heterogeneidad de los estudios deben tomarse con cautela debido al amplio IC 95% de la inconsistencia.

En este último caso, los resultados pueden compararse también con los criterios definidos por Terwee et al (192). Estos autores defienden que la validez de constructo de un cuestionario que mide la AFMV debe obtener un coeficiente de correlación de al menos 0,5 por comparación con los registros de AFMV medidos por acelerometría triaxial. Así mismo, Pols et al. argumentan que la correlación esperada entre un cuestionario de AF y un instrumento comparador apropiado debe oscilar entre 0,5 y 0,7 (177). Sin embargo, los resultados del meta-análisis mostraron que el PAQ-C tampoco cumple con los criterios de suficiente validez de constructo según estos autores.

No se ha encontrado en la literatura científica ninguna otra revisión sistemática que agrupe los resultados sobre un mismo cuestionario de AF, por lo que no podemos cotejar nuestros resultados con los informados por otros autores. Sin embargo, los estudios individuales que valoran la validez de constructo de otros métodos indirectos muestran resultados similares a los reportados en esa revisión. Por ejemplo, la participación en AFMV estimada a

través de cuestionarios y los recuentos de AFMV de acelerómetros presentan correlaciones de entre 0,13 - 0,44 para la versión china del CLASS (299); de 0,19 y 0,38 para las estimaciones de más de 3 y 6 METs en el *Previous Day Physical Activity Recall* (PDPAR) (321); de 0,44 para el cuestionario del SAYCARE (297); o de 0,29 para *Canadian Health Measures Survey* (CHMS) (322).

En relación al procesamiento de los datos de los acelerómetros, cabe destacar la variedad en los enfoques utilizados. Por un lado, para definir un período de medición válido hay autores que determinan como criterio un mínimo de 5 días (208), de 4 días (262,273,278) o de 3 días (282), mientras que otros no lo especifican (280,287). Así mismo, algunos autores establecen un mínimo de horas de registros diarios para considerar un día válido, por ejemplo, 10 horas (278,282), 10 horas para los días entre semana y 8 horas para el fin de semana (273), 9 horas (280) u 8 horas (262), mientras que otros no lo aclaran (208,287). En relación al algoritmo para detectar el tiempo que no se ha llevado el acelerómetro, los estudios que lo definen varían de 10 minutos consecutivos de ceros (278), 20 minutos consecutivos de ceros (262), 60 minutos consecutivos de ceros (280,282) o 60 minutos consecutivos de ceros con tolerancia de 2 minutos (273). Por otro lado, la elección de la época para agrupar los datos brutos difiere entre intervalos de 1 segundo (273), de 5 segundos (262) o de 15 segundos (278,280,282,287). La frecuencia de muestreo solo se especifica en dos de los estudios y se establece en 30 Hz (273,287). Por último, la mayor parte de los estudios que describen el algoritmo para clasificar los recuentos del acelerómetro según los niveles de intensidad (262,273,280,282) utilizan la propuesta de Evenson (323). Solo dos estudios emplean referencias diferentes: Guedes et al. (278) aplican los valores de Fredson (324) y Martín-Bello et al. (287) utilizan los puntos de corte de Schaefer (325).

3.5.6. Puntos fuertes y limitaciones

Un punto fuerte de esta revisión es el seguimiento de una metodología especialmente diseñada para revisiones sistemáticas de propiedades de medición. Por un lado, la evaluación separada de la calidad metodológica de los estudios incluidos y la calificación de los resultados proporciona transparencia en las conclusiones sobre la calidad global de la evidencia. Por otro lado, todos los procedimientos de la revisión fueron realizados por dos investigadores independientes, lo que minimizó la posibilidad de sesgo.

Las limitaciones en cuanto a los resultados obtenidos en esta revisión hacen referencia a la existencia de un único estudio para el EY-PAQ, la poca información sobre el desarrollo y validez de contenido de los cuestionarios incluidos, la falta de datos acerca de la diferencia mínima significativa, la ausencia de estudios que exploren la capacidad de respuesta y la baja calidad de los reportes incluidos.

3.5.7. Consideraciones generales

Los resultados de este estudio demuestran que se necesita más investigación de alta calidad para examinar las propiedades de medición de los cuestionarios del grupo IPAQ para preescolares y escolares.

En primer lugar, para el EY-PAQ solo se dispone del artículo de validez original que ofrece información acerca de la fiabilidad y validez de constructo. Por ello, se recomienda ampliar el estudio de estas y otras propiedades de medición para corroborar si los resultados reportados por los autores son concluyentes.

En segundo lugar, para el PAQ-C se dispone de un mayor número de estudios que analizan sus propiedades de medición. Por un lado, los resultados de fiabilidad se caracterizan por su suficiencia, aunque con protocolos de

medición muy dispares que denotan una alta inconsistencia y dificultan emitir conclusiones con un alto nivel de evidencia. Por otro lado, los estudios del error de medición son escasos y no pueden compararse según los criterios para buenas propiedades de medición debido al desconocimiento de la diferencia mínima significativa del cuestionario. En relación a la validez de constructo, la diversidad de enfoques planteados dificulta establecer una conclusión sólida. En general, la calidad de la evidencia aumenta conforme se utilizan instrumentos de comparación directos que miden el mismo constructo que el cuestionario y pueden aportar un nivel de evidencia 1. Desafortunadamente, los resultados muestran una pobre correlación entre métodos.

En tercer lugar, no se han encontrado estudios que exploren la capacidad de respuesta en ninguno de los cuestionarios incluidos, por lo que se cuestiona su validez como instrumentos para detectar cambios en estudios de intervención.

En cuarto lugar, existe muy poca información acerca de los procedimientos de obtención y reducción de ítems, así como de los procesos de consulta a expertos y a muestras representativas de la población diana. Al tratarse de cuestionarios en los que las preguntas son indicadores del constructo, la falta de evidencia acerca de la validez de contenido plantea si los ítems son capaces de recoger de forma íntegra todos los aspectos claves del constructo o si, por el contrario, no son lo suficientemente exhaustivos y relevantes para la población. Aún así, cabe mencionar que el hecho de que las propiedades de medición analizadas hayan recibido malas calificaciones no significa necesariamente que los cuestionarios no sean adecuados, sino que se ha de ampliar la evidencia con metodologías de alta calidad.

Por último, no se ha detectado ningún cuestionario que cubra el rango de edad de 5 a 7 años, por lo que se propone desarrollar y validar un cuestionario IPAQ que cubra este vacío.

4. Desarrollo y propiedades de medición del
Physical Activity Questionnaire - Young Children
(PAQ-YC)

4.1. Hipótesis general

- El Physical Activity Questionnaire – Young Children (PAQ-YC) es un cuestionario válido y fiable para medir el nivel de actividad física en niños con edades comprendidas entre 5 y 7 años.

4.2. Hipótesis específicas

- El Physical Activity Questionnaire – Young Children (PAQ-YC) muestra una validez de contenido adecuada en términos de validez aparente, relevancia, exhaustividad y comprensibilidad de los ítems.
- No existen efectos techo ni suelo en la puntuación del Physical Activity Questionnaire – Young Children (PAQ-YC).
- La fiabilidad test-retest del Physical Activity Questionnaire – Young Children (PAQ-YC) muestra índices de concordancia intraobservador excelentes.
- La validez de constructo del Physical Activity Questionnaire – Young Children (PAQ-YC) mediante comparación de métodos con acelerometría Actigraph wGT3X-BT (ActiGraph Corp., Pensacola, FL) presenta resultados de correlación suficientes.

4.3. Objetivos generales

- Diseñar el Physical Activity Questionnaire – Young Children (PAQ-YC) para valorar el nivel de actividad física en niños con edades comprendidas entre 5 y 7 años.
- Analizar la validez de contenido, la fiabilidad, el error de medición y la validez de constructo del Physical Activity Questionnaire – Young Children (PAQ-YC).

4.4. Objetivos específicos

- Diseñar el Physical Activity Questionnaire – Young Children (PAQ-YC) en términos de relevancia y exhaustividad de los ítems a través de juicio de expertos usando el método Delphi.
- Desarrollar la validez de contenido en relación a la relevancia, exhaustividad y comprensibilidad de los ítems del Physical Activity Questionnaire – Young Children (PAQ-YC) mediante un proceso de entrevistas cognitivas en una muestra representativa de la población diana.
- Describir el sistema de puntuación del Physical Activity Questionnaire – Young Children (PAQ-YC).
- Estudiar la presencia de efectos techo y suelo en la puntuación del Physical Activity Questionnaire – Young Children (PAQ-YC).
- Analizar la estabilidad y repetitividad del Physical Activity Questionnaire – Young Children (PAQ-YC) a través del estudio de la fiabilidad test-retest.
- Evaluar la validez de constructo del Physical Activity Questionnaire – Young Children (PAQ-YC) mediante comparación de métodos con acelerometría Actigraph wGT3X-BT (ActiGraph Corp., Pensacola, FL).

4.5. Material y métodos

El estudio de desarrollo y propiedades de medición del Physical Activity Questionnaire – Young Children (PAQ-YC) se realizó siguiendo las directrices de construcción y análisis de la guía COSMIN (238).

4.5.1. Desarrollo y validez de contenido

El desarrollo del PAQ-YC se basó en el marco conceptual definido por Pettee Gabriel KK et al. (20), con sus respectivas adaptaciones a la población infantil. Los fundamentos teóricos de este modelo se han descrito ampliamente a lo largo de la introducción de esta tesis, especialmente en los capítulos 1.2, 1.2.3 y 1.2.4.

4.5.1.1. Descripción del constructo

El PAQ-YC tiene por objetivo medir el nivel de AF total practicado por un niño/a sano de entre 5 y 7 años en una semana habitual durante el período lectivo. Incluye las actividades de intensidad ligera a vigorosa realizadas en los dominios escuela (Educación Física u otras iniciativas similares, recreo y pausa de mediodía), transporte y tiempo libre (actividades extraescolares, juego activo en casa e instalaciones interiores y juego en exteriores) durante los últimos 7 días, entendiendo que son representativos de una “semana tipo” siempre y cuando no haya ocurrido ningún evento extraordinario (por ejemplo, una enfermedad) que haya impedido la realización de las actividades habituales. Así mismo, a modo informativo, se incluyen específicamente dos cuestiones relacionadas con el CS durante el tiempo libre, al comprender actividades muy extendidas entre la población infantil. Por tanto, además de medir la AF física habitual, el PAQ-YC pretende ofrecer una estimación del porcentaje de tiempo dedicado a AF y a CS dentro del tiempo de ocio.

El diseño del PAQ-YC se ajustó a su aplicación prevista con fines discriminativos, con la intención de que los resultados del cuestionario sirvan para identificar sujetos suficientemente activos y sujetos inactivos, según las recomendaciones internacionales con respecto al volumen acumulado de AF (3).

Por otro lado, se diseñó para ser utilizado tanto en estudios de investigación (estudios de prevalencia a gran escala y estudios con pequeñas muestras) como en la práctica clínica y académica, es decir, no está supeditado a su aplicación en un entorno específico. El objetivo fue crear un instrumento viable en la medición de la AF a todos los niveles y para todos los profesionales.

Por último, el modo de administración se solventó creando una versión del cuestionario en papel y otra en formato online, para adaptarse a las preferencias de los respondedores. En su construcción, se tuvo en cuenta el tiempo de administración para que no solicitara la atención de los respondedores por más de 15 minutos y se redujo el coste de tiempo unificando en una única versión las respuestas de los padres y los niños/as en lugar de diseñar dos versiones por separado que posteriormente habría que compendiar.

4.5.1.2. Selección de los ítems

El diseño del cuestionario PAQ-YC fue precedido de una revisión exhaustiva de la literatura que permitió obtener un banco de ítems susceptibles de ser incluidos en la versión final. Por un lado, se llevó a cabo una lectura crítica y análisis de los cuestionarios dirigidos a la franja de edad de 5 a 7 años incluidos en las revisiones sistemáticas sobre propiedades de medición publicadas hasta el momento (7-13). Por otro lado, se realizó una revisión sistemática para conocer las características de los cuestionarios del grupo

IPAQ para preescolares y escolares (capítulo 2). La construcción del PAQ-YC tomó como referencia las versiones publicadas del grupo IPAQ para otras franjas de edad en pediatría (224,274), así como las indicaciones de la literatura científica en relación al tipo de AF practicada habitualmente entre los niños de 5 a 7 años (capítulo 1.2.2). Se utilizaron como modelos la versión del *Physical Activity Questionnaire – Children* (PAQ-C) validada al español por Manchola-González et al. para niños de 8-14 años (274) y la versión del *Early Year – Physical Activity Questionnaire* (EY-PAQ) validada por Bingham et al. en Reino Unido para menores de 5 años (224). Se consideró identificar el cuestionario con el nombre *Physical Activity Questionnaire – Young Children* (PAQ-YC), para hacer hincapié en el rango de edad al que va dirigido.

A continuación, se describieron las opciones de respuesta disponibles para cada ítem. Para las preguntas que recogen información sobre el tiempo libre y el transporte, se acordó utilizar un formato de respuesta en forma de cuadrícula con intervalos de tiempo predefinidos cuyo objetivo es reunir información detallada, en términos de duración y frecuencia, de aquellas actividades que están poco estandarizadas. Para los ítems que se refieren a la escuela, se consideró utilizar una escala con tres opciones de respuesta en las que el encuestado debe marcar la opción más afín a la conducta habitual (Figura 26).

El banco de ítems obtenido a través de la revisión de la literatura científica, así como los formatos para las opciones de respuesta fueron discutidos por el equipo de investigación hasta que finalmente se creó una primera versión del cuestionario a través de un proceso de reducción de los ítems de mayor de interés.

Formato tipo cuadrícula. Dominios tiempo libre y transporte:

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

Formato tipo escala con tres opciones. Dominio escuela:

8. ¿Su hijo/a realiza habitualmente las clases de Educación Física en la escuela?

- Mi hijo/a no hace Educación Física
- Mi hijo/a solo hace Educación Física cuando hay que hacer poco esfuerzo físico*
- Mi hijo/a siempre hace Educación Física

Figura 26. Formato de las opciones de respuesta del PAQ-YC

Con respecto a la estructura del cuestionario, se decidió diferenciar dos bloques para hacer una clara división entre las preguntas que el padre/madre pueden responder de forma autónoma (bloque 1) y las preguntas que requieren la colaboración del niño y/o el profesor para confirmar las respuestas en el dominio escuela, donde el padre/madre no está presente. Así mismo, se consideró añadir un subtítulo a cada conjunto de ítems según el dominio y/o el tipo de actividad al que hacen referencia.

4.5.1.3. Elaboración del sistema de puntuación

El algoritmo planteado se inspira en el protocolo de puntuación del IPAQ para adultos, disponible en el sitio web (264). El sistema de puntuación se elaboró una vez se obtuvo la versión final del PAQ-YC. Por razones prácticas y para facilitar su comprensión, se avanzará la versión definitiva de los ítems.

Para los ítems que se refieren a la AF durante el tiempo de ocio (ítems 1, 2 y 3) y el ítem que pregunta por el transporte activo (ítem 4) (Figura 27), se asignó un valor MET estimado utilizando los valores de referencia del compendio para niños entre 6 y 9 años publicados por Butte et al. en 2018 (117). Pese a que la edad del compendio difiere ligeramente de la población diana del PAQ-YC, no se consideró hacer ningún ajuste en base a la TMR ya que probablemente habría generado una fuente de error de mayor magnitud al estandarizar el peso y estatura medias de los niños entre 5 y 7 años (109). En la Tabla 19 se presentan las actividades para cada uno de los ítems del 1 al 4 con sus respectivos METs individuales extraídos del compendio infantil y el valor medio por categoría.

1. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha hecho actividades dirigidas por monitor/entrenador en horario extraescolar que requieran estar activo (por ejemplo, practicando deportes, danza...)? No tenga en cuenta en esta pregunta el tiempo en la escuela, como el recreo o la Educación Física.

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

2. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha jugado activamente en casa (por ejemplo, bailando, corriendo, empujando juguetes, jugando a videojuegos activos como la Nintendo Wii...) o en instalaciones de juego interiores (por ejemplo, en una ludoteca, en una piscina de bolas...)?

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

3. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha jugado activamente en exteriores (por ejemplo, en el jardín, en el parque...)? No tenga en cuenta las actividades extraescolares dirigidas que se han incluido en la pregunta 1 ni el tiempo de recreo en la escuela.

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

4. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a se ha desplazado de un sitio a otro de forma activa (por ejemplo, caminando, en bicicleta, en patinete...) sin utilizar ningún medio de transporte (ni coche, ni autobús, ni metro...)? Incluya en esta pregunta si el niño/a va caminando a la escuela y otros posibles desplazamientos (por ejemplo, ir de compras, ir a las actividades extraescolares...).

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

Figura 27. Ítems 1, 2, 3, y 4 del PAQ-YC

Tabla 19. Equivalencias METs para los ítems 1, 2, 3 y 4 del PAQ-YC

Ítem 1		Ítem 2		Ítem 3		Ítem 4	
AF	METs	AF	METs	AF	METs	AF	METs
Baloncesto - juego mini	4,9	Videojuego - Kinect	3,1	Juegos de pelota	6	Bicicleta (RP)	4,6
Atrapar/lanzar pelota	4,1	Videojuego - Wii	2,3	Juegos tipo brilé	5,8	Monopatín	5,7
Golg - juego mini	4	Videojuego - conducir	2	Juego libre	5,7	Patinete	4,9
Balonmano	5,4	Videojuego - disparar	2,3	Rayuela	5,8	Hoveboard	1,8
Hockey - juego mini	3,8	Arcade - fútbolín	1,9	Saltar a la comba	6,3	Caminar (RP)	3,6
Patinaje	5,2	Juego de mesa - de pie	2	Juegos - miscelánea	6,4		
Ski	5,6	Pintar - de pie	1,8			Media	4,1
Fútbol - conos	5,4	Cantar - de pie	1,8	Media	6,0	SD	1,5
Fútbol - juego	7,7	Apilar objetos	1,6	SD	0,3		
Tenis de mesa	4,2						
Tenis - juego	6,1	Media	2,1				
Vóleibol	5	SD	0,4				
Aerobic - danza	3,6						
Natación (RP)	9,5						
Media	5,3						
SD	1,6						

RP: ritmo propio, AF: actividad física, METs: equivalentes metabólicos, SD: desviación estándar

Para los ítems 5 y 6 no se calcularon los METs asociados ya que la categoría de actividades que recogen encaja dentro de los CS, todas ellas reconocidas con un valor inferior a 1,5 METs (Figura 28). Además, estos ítems no forman parte de la puntuación total del cuestionario, al entenderlos como un comportamiento diferente a la AF. Sin embargo, los minutos semanales informados en estas dos preguntas se presentan como un resultado secundario de tipo informativo que sirve para calcular el porcentaje estimado de tiempo dedicado a AF y CS discrecional.

5. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha estado sentado jugando con juguetes o sin moverse haciendo actividades como colorear, dibujar, hacer manualidades, hacer los deberes...? No tenga en cuenta las horas en la escuela ni el tiempo con dispositivos electrónicos (por ejemplo, la TV, el ordenador, los videojuegos...).

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

6. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha estado sentado o sin moverse haciendo actividades como ver la TV, hacer los deberes con ordenador o Tablet, jugar con un videojuego, jugar al ordenador, móvil o Tablet? No tenga en cuenta las horas en la escuela ni los videojuegos activos como la Nintendo Wii.

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

Figura 28. Ítems 5 y 6 del PAQ-YC

Las respuestas del cuestionario para estos seis primeros ítems se presentan como rangos de duración para cada día. Como probablemente el tiempo dedicado a cada categoría de actividad no se reparta de forma equitativa durante los 7 días de la semana, se decidió calcular el promedio de minutos para trabajar con un único dato por ítem. Para saber qué valor tomar dentro del rango se replicó la metodología del EY-PAQ (224), es decir, se utilizó un enfoque pragmático con respecto a la duración informada. En los ítems que preguntan sobre comportamientos activos (ítems 1, 2, 3 y 4) se consideró

coger el valor inferior del rango, mientras que para los ítems 5 y 6, que miden CS, se decidió tomar el valor superior.

Para calcular la puntuación del primer bloque, se tuvieron en cuenta solo los ítems 1, 2, 3 y 4. A cada ítem se le asignó un valor en una escala Likert de 0-5 en función del volumen acumulado de AF expresado mediante el promedio de minutos diarios y el coste MET estimado asociado a las actividades de la categoría. Los criterios de puntuación se elaboraron de acuerdo con las recomendaciones de AF que aconsejan que los niños entre 5 y 17 años practiquen al menos 60 minutos de AFMV al día (3). Por ejemplo, se asignó un valor de 3 puntos si el promedio de minutos diarios combinado con el valor MET para cada ítem (Tabla 20) alcanzaba un mínimo de 180 MET-min/día, lo cual equivaldría a 60 minutos de AF de intensidad moderada (cumplimiento mínimo de las recomendaciones de AF) (3).

Tabla 20. Equivalencias de puntuación para los ítems 1, 2, 3 y 4 según los MET-min/día

Puntuación	MET-min/día	Min/día AF moderada	Min/día Ítem 1	Min/día Ítem 2	Min/día Ítem 3	Min/día Ítem 4
0	≤ 29,99	≤ 9,99	≤ 5,65	≤ 14,28	≤ 5,00	≤ 7,31
1	30 - 89,99	10 - 29,99	5,66 - 16,98	14,29 - 42,85	5,01 - 15,00	7,32 - 21,95
2	90 - 179,99	30 - 59,99	16,99 - 33,96	42,86 - 85,71	15,01 - 30,00	21,96 - 43,90
3	180 - 269,99	60 - 89,99	33,97 - 50,94	85,72 - 128,57	30,01 - 45,00	43,91 - 65,85
4	270 - 359,99	90 - 119,99	50,95 - 67,92	128,58 - 171,42	45,01 - 60,00	65,86 - 87,80
5	≥ 360	≥ 120	≥ 67,93	≥ 171,43	≥ 60,01	≥ 87,81

Para establecer el sistema de puntuación de los ítems del bloque 2, de nuevo se estudiaron los costes energéticos de las actividades que recoge cada uno de ellos, excepto para la pregunta 7 que funciona como un factor de corrección aplicado al ítem 8.

Tal y como se ha explicado en el capítulo 1.2.2.2, el tiempo dedicado a la Educación Física a la semana suele variar de 1 a 3 horas entre centros, aunque el 44,2% declara invertir 180 min o más (49). Por este motivo, se consideró que estandarizar una misma puntuación para el ítem 8 podría conllevar un error de medición importante. En este contexto, el ítem 7 se diseñó como un ítem no puntuable en sí mismo que funciona como factor de corrección aplicable al ítem 8 según las horas de Educación Física practicadas en el centro escolar. Con respecto al coste energético, se consideró que, por la similitud de las actividades recogidas en cada categoría, la intensidad en METs asociada al ítem 8 equivaldría al ítem 1 (estimación de 5,3 METs) si se participa con normalidad en la asignatura, mientras que se correspondería con el valor MET asignado al ítem 2 (estimación de 2,1 METs) si se participa en la Educación Física solo cuando se requiere poco esfuerzo físico. Por último, si se informa que el niño/a nunca hace Educación Física, el comportamiento que refleja el ítem encaja dentro de una conducta sedentaria, por lo que directamente se asignaría un valor de 0 puntos, independientemente de la respuesta a la pregunta 7. De este modo, la puntuación del ítem 8 se establece tal y como se presenta en la Tabla 21.

Del mismo modo, para los ítems 9 y 10, se consideró que el coste energético podría ser equiparable a la intensidad en METs asociada a los ítems 2 y 3 dependiendo de si se realiza poco esfuerzo físico (estimación de 2,1 METs) o si se participa de forma enérgica todo el tiempo (estimación de 6,0 METs). Por otro lado, tal y como se ha explicado en el capítulo 1.2.2.2, la duración del recreo es de 30 minutos al día, mientras que la duración de la pausa para comer es de aproximadamente 2 horas. Con respecto al descanso a mediodía, se consideró que la oportunidad para la AF ocurre durante aproximadamente 30 minutos, teniendo en cuenta que el resto del tiempo se consume en los desplazamientos, la comida, los hábitos higiénicos, etc. De nuevo, la puntuación de los ítems 9 y 10 depende de la opción de respuesta marcada en el cuestionario y se establece tal y como se muestra en la Tabla 22.

Tabla 21. Puntuación del ítem 8 con factor de corrección del ítem 7.

Respuesta ítem 8	Respuesta ítem 7		Puntuación ítem 8
	Horas/sem	MET-min/día	
“Mi hijo/a no hace Educación Física” 0			0
“ Mi hijo/a solo hace Educación Física cuando hay que hacer poco esfuerzo físico” 1	1	18,00	0
	1,5	27,00	0
	2	36,00	1
	2,5	45,00	1
	3	54,00	1
	3,5	63,00	1
	4	72,00	1
“Mi hijo/a siempre hace Educación Física” 2	1	45,42	1
	1,5	68,10	1
	2	90,84	2
	2,5	113,58	2
	3	136,26	2
	3,5	159,00	2
	4	181,68	3

Tabla 22. Puntuación de los ítems 9 y 10.

Opción de respuesta	MET-min/día	Puntuación
“Estar sentado (hablar, leer, jugar a juegos sin moverse, deberes...)”		0
“ Correr y jugar a juegos que impliquen poco esfuerzo físico”	45,00	1
“Correr y jugar intensamente todo el tiempo”	128,57	2

La puntuación del PAQ-YC con respecto a la AF total se expresa como una medida adimensional que varía de 0 a 27 puntos, siendo los valores más altos indicativos de un mayor nivel de AF. Para calcularla, basta con sumar la puntuación obtenida en los ítems 1, 2, 3, 4, 8 corregido, 9 y 10.

Tal y como se ha avanzado es este capítulo, además de la puntuación total, se puede presentar otro resultado de tipo informativo que muestra una estimación del porcentaje de tiempo dedicado a la AF y el CS discrecional. Para calcular este índice se deben sumar, por un lado, los minutos informados en los ítems 1, 2, 3 y 4 y, por otro, los minutos descritos en las preguntas 5 y 6. A continuación, se calcula la proporción de tiempo de cada comportamiento con respecto a los minutos totales.

4.5.1.4. Consultas a expertos y a muestras de la población diana

Para evaluar la validez de contenido del PAQ-YC en términos de relevancia y exhaustividad de los ítems, se creó una consulta por juicio de expertos mediante el método Delphi. Se envió la propuesta de colaboración por correo electrónico a potenciales expertos relacionados con el ámbito de la AF en pediatría pertenecientes a las ramas de conocimiento de Ciencias de la Actividad Física y el Deporte (CAFE), fisioterapia en pediatría, Educación Infantil y Educación Primaria.

Para valorar la idoneidad de los seleccionados se obtuvo el nivel de experticia de cada uno de ellos calculando el Coeficiente de Conocimiento (Kc) como la media ponderada de la puntuación en cada uno de los ítems de la Tabla 23 donde el peso relativo se determinó en función del grado de influencia. El Kc se calculó en base a la información que el propio experto presentaba del tema de estudio con respecto a la experiencia práctica y los conocimientos teóricos, mediante un proceso de autovaloración en una escala de cierre gráfico [0-10]. Se hizo la criba de aquellos expertos que en el Kc no alcanzaron el nivel crítico exigible de 0,8 (326).

Tabla 23. Ponderación del Coeficiente de Conocimiento (Kc)

Ítems de valoración	Peso
Conocimiento sobre la AF en centros educativos españoles	0,20
Conocimiento sobre la AF en horario extraescolar que practican los niños	0,20
Experiencia práctica en el ámbito de la AF en etapa escolar	0,25
Conocimiento sobre las recomendaciones de AF en etapa escolar	0,25
Habilidad en el manejo de herramientas de valoración del nivel de AF	0,10

Posteriormente, se realizaron las rondas de consulta del cuestionario con los miembros del panel de expertos que obtuvieron el Kc mínimo exigible y estaban dispuestos a colaborar. Se elaboró una consulta online en la que se presentaban las características generales y los ítems que formaban la propuesta del PAQ-YC con el objetivo de generar un debate que permitiera una retroalimentación de manera ordenada. Las rondas de consulta siguieron un proceso iterativo, es decir, una vez obtenidas las respuestas en una primera ronda, las preguntas donde no hubo consenso fueron nuevamente formuladas. El procedimiento se repitió tantas veces como fue necesario hasta llegar a un acuerdo. La identidad de los participantes se mantuvo en el anonimato para evitar que los miembros del panel con mayor experiencia pudieran condicionar las respuestas de los otros.

Se pidió a los expertos que expresaran su grado de acuerdo con cada uno de los ítems, así como con las características generales del cuestionario para evaluar la validez aparente (claridad del contenido introductorio e instrucciones, facilidad de comprensión, claridad y sencillez del formato, adecuación de la división en dos bloques, adecuación del número de preguntas, adecuación del contenido de los ítems). El grado de acuerdo se expresó en una escala de 1 a 5 siendo 1 la puntuación asignada para el valor más bajo posible "totalmente en desacuerdo", 2 para el valor "en desacuerdo", 3 para el valor "indiferente", 4 para el valor "de acuerdo" y 5 para el valor "totalmente de acuerdo". Además, se habilitó un apartado de comentarios

para cada uno de los ítems y características generales con la intención de que, además de expresar su grado de acuerdo, pudieran manifestar propuestas de mejora específicas.

Para cada ronda de preguntas, se calculó la mediana y los cuartiles 1 (Q1) y 3 (Q3) de las puntuaciones obtenidas en cada respuesta. Se estableció como criterio de consenso que el Rango Inter cuartil Relativo (RIR) $((Q3 - Q1/Me) \times 100)$ y/o el Coeficiente de Variación (CV) $((SD/Me) \times 100)$ fueran menores o iguales al 20% (327).

Una vez obtenida la versión consensuada por los expertos, se llevó a cabo un proceso de entrevistas cognitivas en una muestra de padres y madres de niños entre 5 y 7 años para asegurar que todos los ítems eran comprensibles y considerados como relevantes y exhaustivos según la población diana.

En primer lugar, se les entregó a las familias el cuestionario en papel y se les pidió que lo respondieran de forma autónoma sin entablar conversación con el investigador. A continuación, se llevó a cabo una entrevista semiestructurada en la que se preguntó sobre la relevancia, exhaustividad y comprensibilidad de cada uno de los ítems según el orden en el que se muestran en el cuestionario. El enfoque utilizado para interrogar a las familias fue el sondeo verbal. Después de que el participante respondiera el cuestionario, el entrevistador indagó sobre la interpretación de cada uno de los ítems para explorar la comprensibilidad del redactado y el proceso de recuperación de recuerdos para obtener la respuesta. Así mismo, se indagó sobre la opinión de las familias con respecto a las características generales del PAQ-YC en relación a la claridad del contenido introductorio, la claridad y sencillez del formato, la adecuación de la división en dos bloques y el número de preguntas. Por último, se comentaron los temas mencionados espontáneamente por las familias. Se tomaron notas escritas durante todo el proceso para registrar la opinión de las familias, pero no se creyó conveniente realizar grabaciones de la entrevista para no cohibir a los participantes.

4.5.1.5. Prueba de campo

Una vez obtenida la versión final del PAQ-YC se llevó a cabo una prueba de campo para estudiar la distribución de los ítems del cuestionario.

El cálculo del tamaño de la muestra siguió la regla 1:10 para el ratio de observaciones y el número de ítems, estimándose un tamaño de muestra mínimo de 100 participantes (328).

Se incluyeron niños con edades comprendidas entre los 5 y 7 años y sus respectivos padres/madres. Se establecieron como criterios de inclusión estar escolarizados en escuela ordinaria y tener una buena comprensión oral y escrita del español. Los participantes se reclutaron mediante un muestreo por voluntarios en las comunidades de Galicia, Cataluña, Islas Baleares y La Rioja.

El cuestionario fue entregado en papel o formato electrónico (según preferencia) a los padres de los participantes. Se aseguró que en los 7 días previos a la administración del PAQ-YC no hubiera festividades, vacaciones o cualquier otra circunstancia especial que afectase la escolaridad durante los 5 días de la semana. Se recogieron las variables sexo, edad, curso escolar y parentesco del adulto respondedor, así como todas las que se desprenden de cada uno de los ítems del cuestionario.

Al tratarse de un cuestionario construido bajo un modelo formativo, se descartó aplicar métodos estadísticos como el Análisis Factorial Exploratorio (EFA) (201) o la Teoría de Respuesta al Ítem (IRT). Sin embargo, se observó la aportación de cada uno de los ítems a los resultados del PAQ-YC para determinar si alguno de ellos tenía un impacto reducido en la puntuación. En ese caso, el equipo valoró si se podía deducir que su inclusión en el cuestionario era irrelevante para la mayoría de los participantes.

Por último, se analizó la presencia de valores extremos en la puntuación del cuestionario. Se consideró que existía efecto techo o suelo si más del 15% de

los individuos presentaban la máxima o la mínima puntuación, respectivamente (192).

4.5.2. Fiabilidad y error de medición

En una siguiente fase, se analizó la fiabilidad y error de medición del PAQ-YC en una submuestra de los participantes de la prueba de campo.

En el cálculo del tamaño de la muestra para el análisis de fiabilidad se utilizaron las fórmulas desarrolladas en 2012 por Zou et al. (329). Se establecieron como parámetros un nivel de concordancia esperado de 0,85 y no inferior a 0,75 con un nivel de confianza del 5% y una potencia del 80%, estimándose un tamaño de muestra necesario de 78 individuos.

El cuestionario fue entregado en papel o formato electrónico (según preferencia) a los padres de los participantes y se administró en dos ocasiones con un lapso entre 1 y 3 días. Como el cuestionario valora la AF en los últimos 7 días, se pidió que tomaran como referencia la misma semana, sin tener en cuenta el tiempo transcurrido entre el test y el retest. Se aseguró que los participantes se mantuvieran estables entre la primera y la segunda administración y se confirmó que las condiciones de la prueba eran las mismas para ambas mediciones en relación al formato de administración, el entorno y las instrucciones.

Para el análisis estadístico de los datos se utilizó el programa SPSS v24.0 (*IBM Corp. Released 2016. IBM SPSS Statistics for Windows Armonk, NY*). Para la evaluación de la fiabilidad de los ítems que contribuyen a la puntuación del cuestionario se utilizó el coeficiente de Kappa, mientras que para el análisis de fiabilidad de los bloques 1 y 2, de la puntuación total y de los porcentajes de AF y CS discrecional se calculó el ICC con un modelo aleatorio (2,1).

Para determinar el umbral del coeficiente de fiabilidad Kappa se estableció un mínimo de 0,70 (192). Para determinar el umbral de fiabilidad del ICC se siguieron los criterios de Fleiss et al. (330): si coeficiente de fiabilidad $> 0,8$ la fiabilidad se considera excelente; si $0,6 < ICC \leq 0,8$ la fiabilidad se considera buena; si $0,4 < ICC \leq 0,6$ la fiabilidad se considera moderada y si $ICC \leq 0,4$ la fiabilidad se considera débil o pobre.

El error de medición se expresó como el SEM en todos los casos. Así mismo, para los ítems que contribuyen a la puntuación del cuestionario, la puntuación del bloque 1, la puntuación del bloque 2 y la puntuación total se calculó el PoA entre la primera y la segunda administración.

4.5.3. Validez de constructo

La última fase del estudio de propiedades de medición del PAQ-YC consistió en el análisis de la validez de constructo por comparación de métodos en una submuestra de los participantes de la prueba de campo.

Se utilizó el acelerómetro triaxial ActiGraph wGT3x-BT (ActiGraph Corp., Pensacola, FL) para evaluar los movimientos corporales en las horas de vigilia durante siete días consecutivos. El modelo ActiGraph wGT3X-BT (ActiGraph Corp., Pensacola, FL) captura y registra información de actividad de alta resolución utilizando un acelerómetro MEMS de estado sólido de 3 ejes con un rango dinámico de ± 8 G. Los datos de aceleración son muestreados por un convertidor analógico digital de 12 bits a velocidades especificadas por el usuario que van desde 30 a 100 Hz y se almacenan sin procesar en una memoria flash no volátil. Los algoritmos utilizados están patentados y han sido validados por miembros de la comunidad científica de investigación académica. Cada unidad se calibra como parte del proceso de fabricación estándar de ActiGraph, LLC y se mantiene durante la vida útil del producto (331).

Se instruyó a los participantes en la colocación del dispositivo en la cintura, a la altura de la línea media clavicular izquierda. Se les pidió que lo retiraran para actividades bajo el agua y para dormir (por motivos de confort del usuario). Para ello, las familias que participaban en el estudio recibieron información gráfica a través un vídeo diseñado para explicar el objetivo del estudio, así como los puntos claves en el uso del dispositivo. Los acelerómetros fueron programados para registrar datos a una frecuencia de 50 Hz.

El PAQ-YC fue entregado en papel o formato electrónico (según preferencia) a los padres de los participantes una vez finalizado el período de recogida de datos del acelerómetro. Se pidió que respondieran a las preguntas haciendo referencia a los mismos siete días en los que se había llevado puesto el dispositivo.

El cálculo del tamaño de la muestra fue de 30 individuos, teniendo en cuenta un coeficiente de correlación de 0,5 (192) y aceptando un riesgo alfa de 0,05 y un riesgo beta de 0,2 en un contraste bilateral. Se estimó una tasa de pérdidas de del 25% por lo que se consideró reclutar un mínimo de 39 participantes.

Los datos brutos de aceleración se agruparon en intervalos de época de 10 segundos utilizando el software ActiLife. El tiempo sin uso se estableció mediante los datos de aceleración sin procesar utilizando el enfoque de Choi et al. (332) modificado. Se definió el tiempo sin uso como los períodos consecutivos de ceros durante 60 minutos con tolerancia a valores distintos durante 2 minutos para permitir la detección de artefactos siempre y cuando los datos de los 30 minutos previos o posteriores fueran ceros consecutivos. Se consideró un día válido si el tiempo de uso del dispositivo superaba las 10 horas. Se incluyeron los datos de aquellos participantes con un mínimo de 6 días válidos. Los resultados de cada participante se filtraron de forma individual para establecer los períodos de tiempo discrecional. Se diseñaron

dos filtros para definir las horas libres de los días entre semana según si el horario escolar del participante era continuo (desde las 14:00 hasta las 23:59 horas) o partido (desde las 16:30 hasta las 23:59 horas) y se combinaron con otro filtro que incluía los fines de semana. Para clasificar los resultados según los niveles de intensidad se utilizaron los puntos de corte definidos por Evenson et al. en niños de 5 a 8 años (323). Los datos se presentaron como medidas de magnitud vectorial ($VM = \sqrt{(Axis\ 1)^2 + (Axis\ 2)^2 + (Axis\ 3)^2}$) para recoger los registros de los tres ejes del dispositivo.

Para el análisis estadístico de la concordancia entre el PAQ-YC y los registros del acelerómetro se utilizó el programa SPSS v24.0 (*IBM Corp. Released 2016. IBM SPSS Statistics for Windows Armonk, NY*). Los resultados se presentaron como coeficientes de correlación de Pearson o Spearman en función de si los datos seguían o no una distribución normal. Se comprobó si los datos presentaban una distribución normal con el test de Shapiro-Wilk. Así mismo, se muestran los gráficos de dispersión entre las variables analizadas para la validez de constructo.

Se consideró que existía validez de constructo perfecta si el coeficiente de correlación entre ambas medidas era igual a 1. Para los demás valores, se determinó que la validez de constructo era excelente si se encontraba entre 0,9 y 1; fuerte entre 0,75 y 0,9; moderada entre 0,5 y 0,75; y débil entre 0,1 y 0,5. Se estableció como criterio de suficiencia un mínimo de correlación de 0,5 en base a las recomendaciones de Terwee et al. (192) y se estudió si los resultados podían proporcionar un nivel de evidencia 1 según los criterios de Hidding et al. (13).

4.5.4. Consideraciones éticas

Las fases de desarrollo del cuestionario se realizaron de acuerdo con las consideraciones éticas de la Declaración de Helsinki. Todos los miembros que

colaboraron fueron informados del objetivo del estudio y del procedimiento que se llevó a cabo. De acuerdo con el comité de ética en experimentación animal y humana de la *Universitat Internacional de Catalunya* y debido a que no se trataba de un estudio experimental, se concluyó que no era necesario realizar el consentimiento informado. Sin embargo, para los estudios de campo, el proyecto se presentó al comité de ética de investigación de la *Universitat Internacional de Catalunya* (Anexo 8.4). Todos los padres de los participantes recibieron información sobre el estudio y dieron su consentimiento para la participación (Anexos 8.5 y 8.6).

Se mantuvo en todo momento la confidencialidad de los participantes cumpliendo la ley de protección de datos 3/2018, de manera que el acceso a cualquier documento o archivo obtenido durante el transcurso del estudio estuvo restringido a toda persona ajena a la investigación. Los datos personales de los participantes se codificaron con un número y en ningún momento se reveló su identidad.

4.6. Resultados

4.6.1. Desarrollo y validez de contenido

La propuesta inicial del PAQ-YC constaba de dos bloques estructurados de forma ordenada para guiar al respondedor a lo largo del cuestionario. El bloque 1 debía ser respondido por los padres y contenía 7 ítems que hacían referencia a los dominios tiempo libre y transporte. El bloque 2 contenía 3 ítems relacionados con la AF en la escuela. Para responder a las preguntas del bloque 2, los padres debían establecer un diálogo con el niño para marcar la respuesta.

Esta primera versión del cuestionario fue sometida a juicio de expertos mediante el método Delphi. La propuesta fue enviada a 48 potenciales expertos, de los cuales se obtuvo respuesta de 34 que rellenaron el formulario de autovaloración. Solo 22 (64,7%) de estos potenciales expertos cumplían con el criterio de experticia requerido y fueron seleccionados para la evaluación del cuestionario. El panel final estuvo formado por 13 expertos (7 hombres, 6 mujeres; edad = 38,85 (11,43), Kc = 8,67 (0,7)) en la primera ronda y 11 expertos (5 hombres, 6 mujeres; edad = 39,91 (11,99); Kc = 8,50 (0,61)) en la segunda ronda. En la primera, ronda se contó con la participación de 3 fisioterapeutas, 2 titulados en CAFE, 2 fisioterapeutas y titulados en CAFE, 4 maestros de Educación Primaria, 1 maestro de Educación Infantil y 1 maestro de Educación Infantil y psicólogo. En la segunda ronda, no se obtuvo respuesta por parte de dos de los maestros de Educación Primaria que habían participado en la primera consulta.

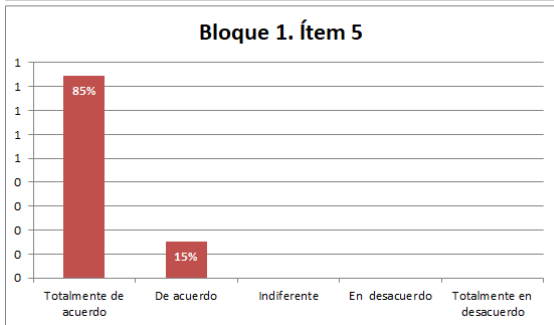
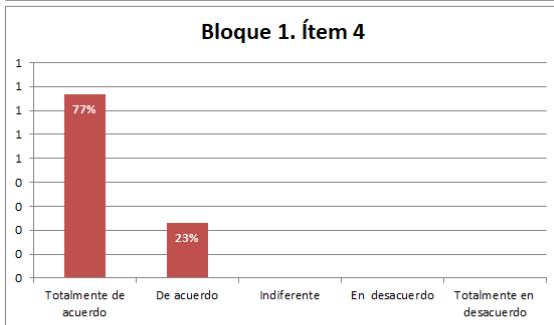
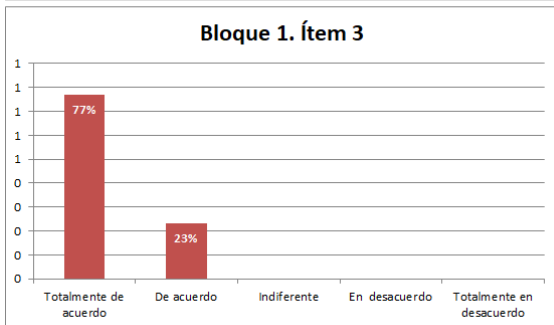
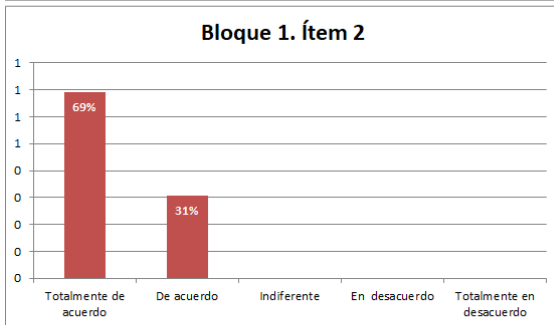
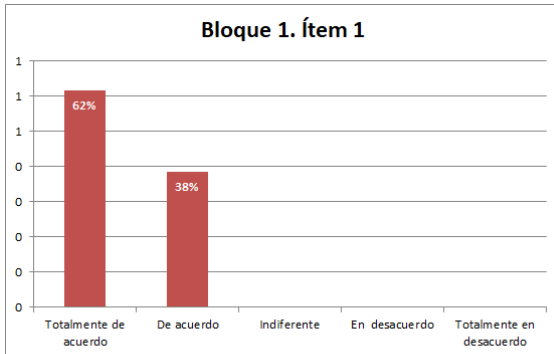
Se necesitaron dos rondas de consulta para alcanzar consenso. En la primera ronda, los expertos estuvieron “totalmente de acuerdo” o “de acuerdo” en los ítems 1-5 del primer bloque, en el ítem que pregunta sobre la representatividad de una semana habitual: “¿Estuvo enfermo su hijo/a en la última semana o algo impidió que hiciera las actividades habituales?” y en las

características generales del cuestionario. En la Figura 29 se muestran los porcentajes de respuesta para cada uno de los ítems y las características generales en las rondas de respuesta 1 y 2. Las preguntas del bloque 2 así como la pregunta 7 del bloque 1 generaron una mayor discrepancia en el panel de expertos mostrando un RIR y un CV mayor del 20% (Tabla 24 y Tabla 25).

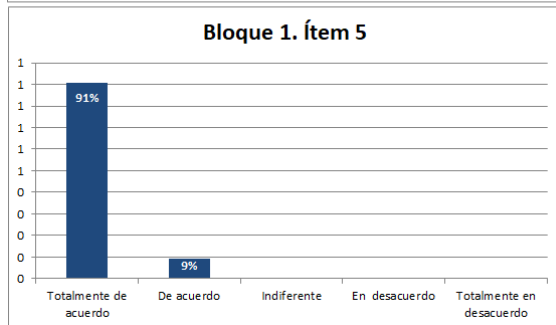
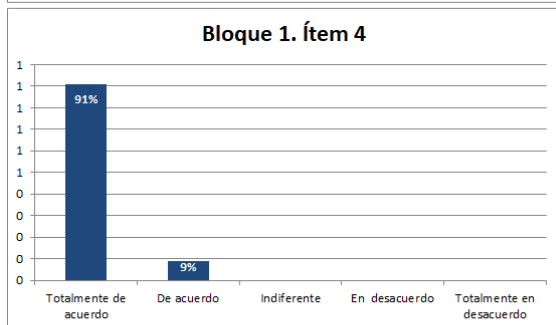
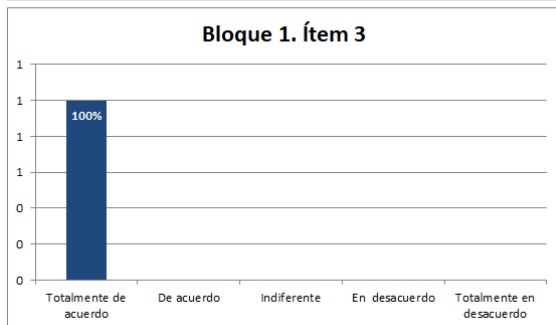
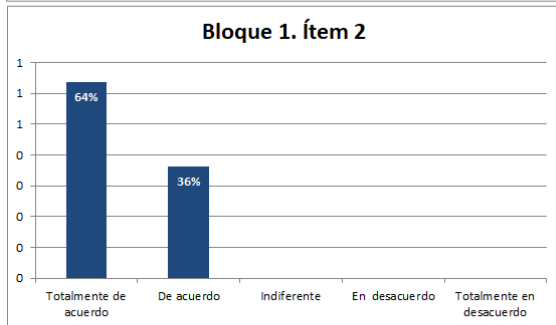
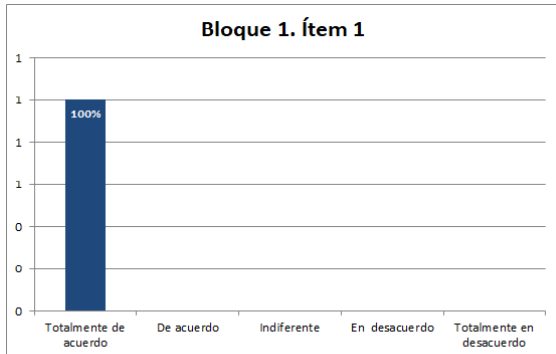
En base a sus recomendaciones, se incluyeron pequeñas modificaciones de redactado (contenido introductorio e ítems del 1 al 6) o modificaciones sustanciales en el enunciado del ítem y/o en las opciones de respuesta (ítems del bloque 2). Así mismo, se decidió eliminar el ítem 7 del bloque 1: *“¿Cuál de las siguientes frases cree usted que describe mejor la última semana de su hijo/a? Lea las 5 frases antes de decidir cuál la describe mejor.”* y añadir un nuevo ítem en el bloque 2: *“¿Cuántas horas de Educación Física o actividades similares hacen a la semana en el curso de su hijo/a? En el curso de mi hijo/a hacen horas y minutos de Educación Física a la semana”*.

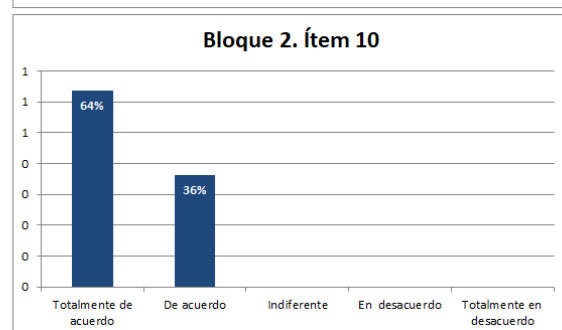
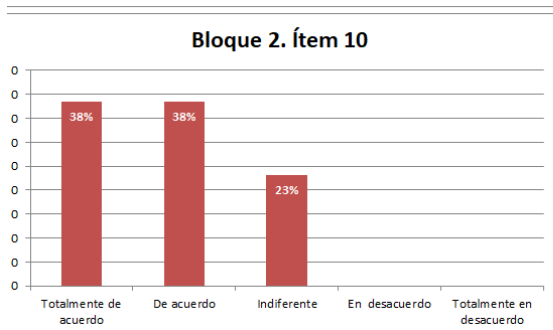
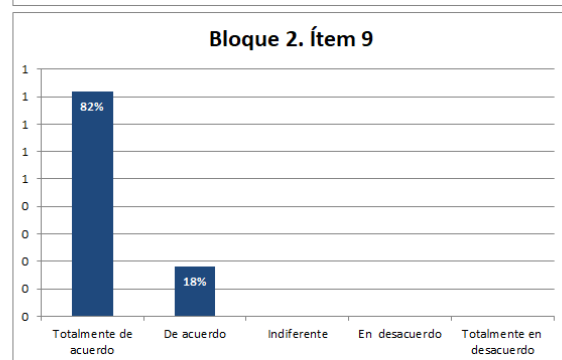
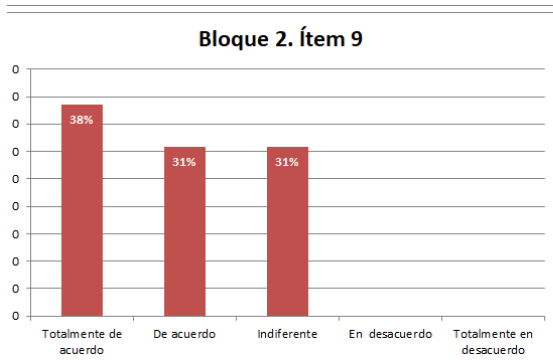
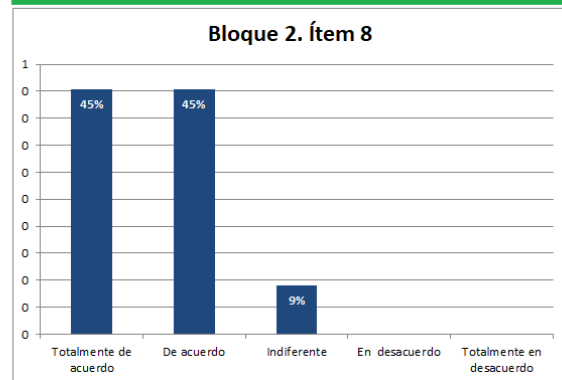
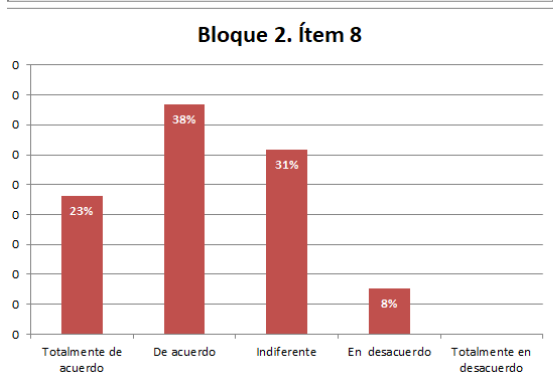
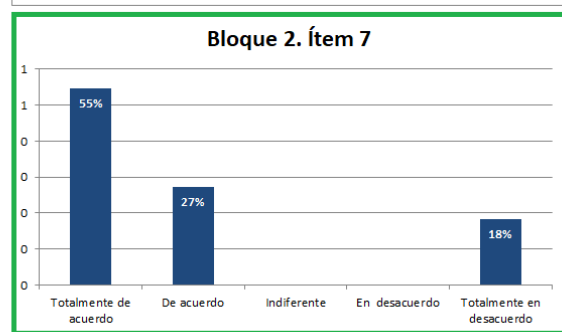
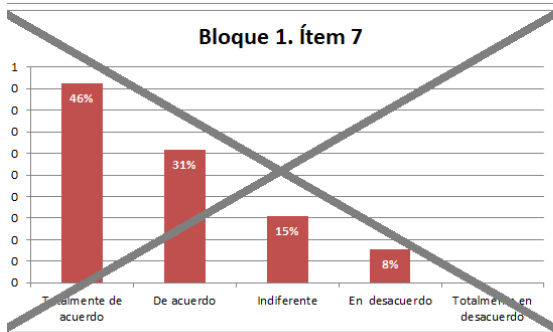
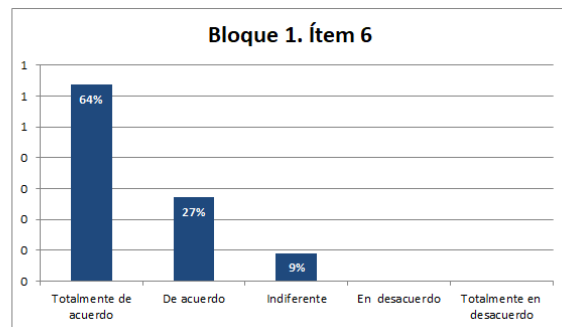
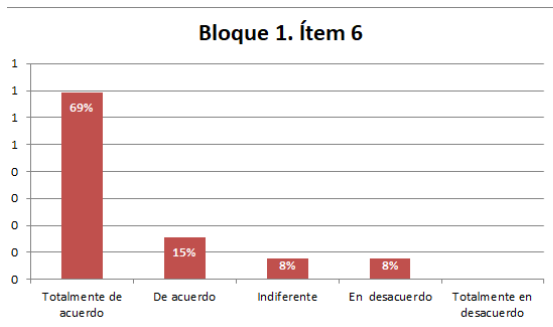
En la ronda 2, los expertos estuvieron “totalmente de acuerdo” o “de acuerdo” en los ítems 1-5 del bloque 1, en los ítems 9 y 10 del bloque 2, en el ítem que pregunta sobre la representatividad de una semana habitual y en las características generales del cuestionario. Cabe destacar que en los ítems 6 y 8 hubo un único experto que no expresó su grado de acuerdo como “totalmente de acuerdo” o “de acuerdo”, mientras que para el ítem 7 fueron 2 expertos quienes disintieron de la opinión general. Sin embargo, en todos los casos pudo considerarse que existía acuerdo ya que se cumplía al menos uno de los criterios establecidos (ítem 6: RIR = 20,00 y CV = 15,45; ítem 7: RIR = 20,00 y CV = 38,73; ítem 8: RIR = 25,00 y CV = 15,45) (Tabla 24 y Tabla 25).

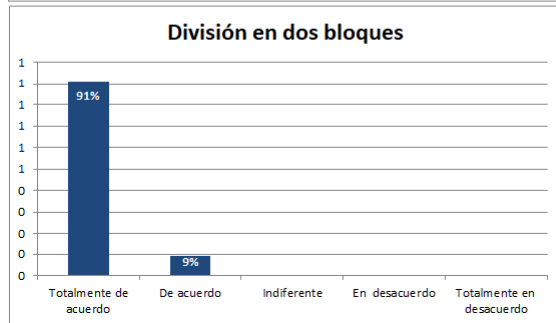
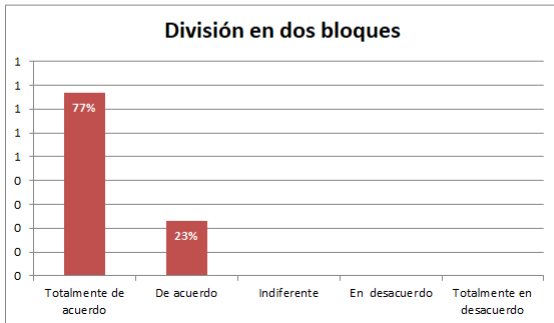
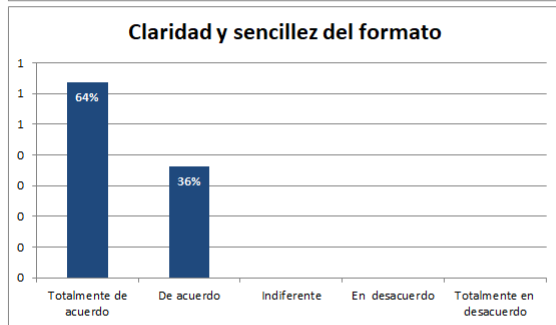
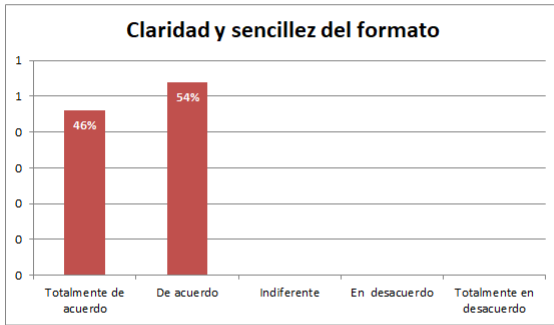
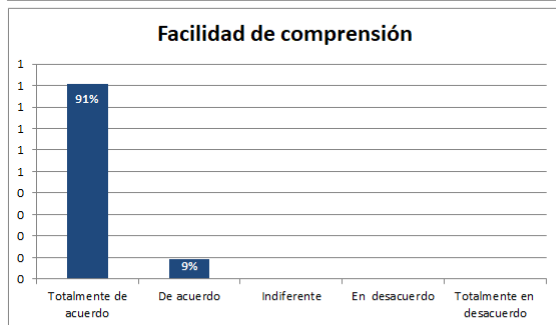
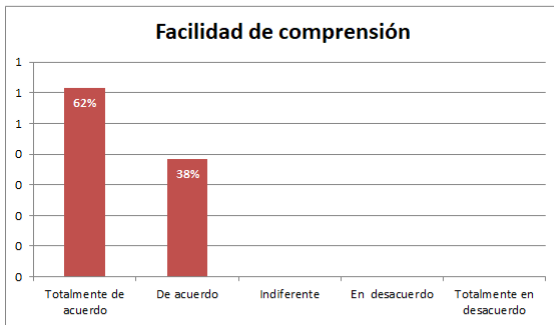
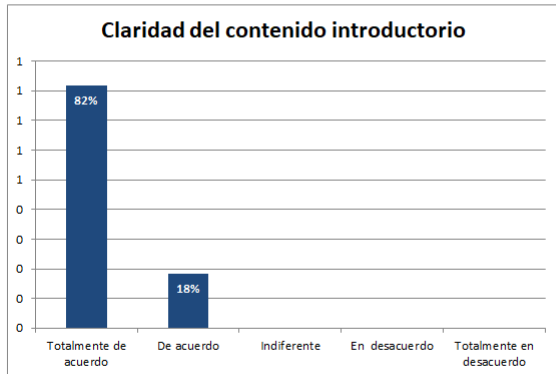
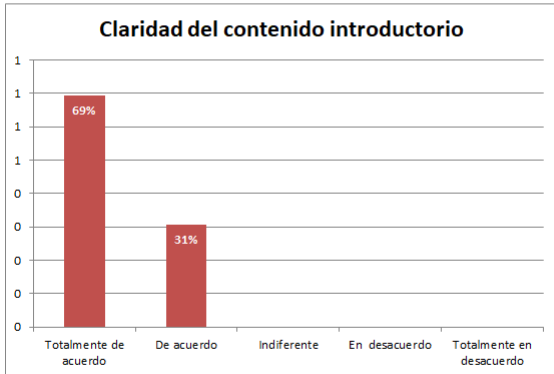
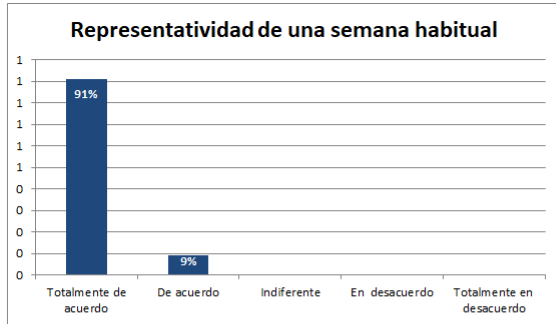
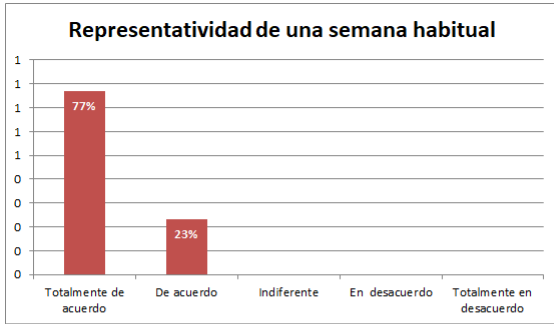
Ronda 1



Ronda 2







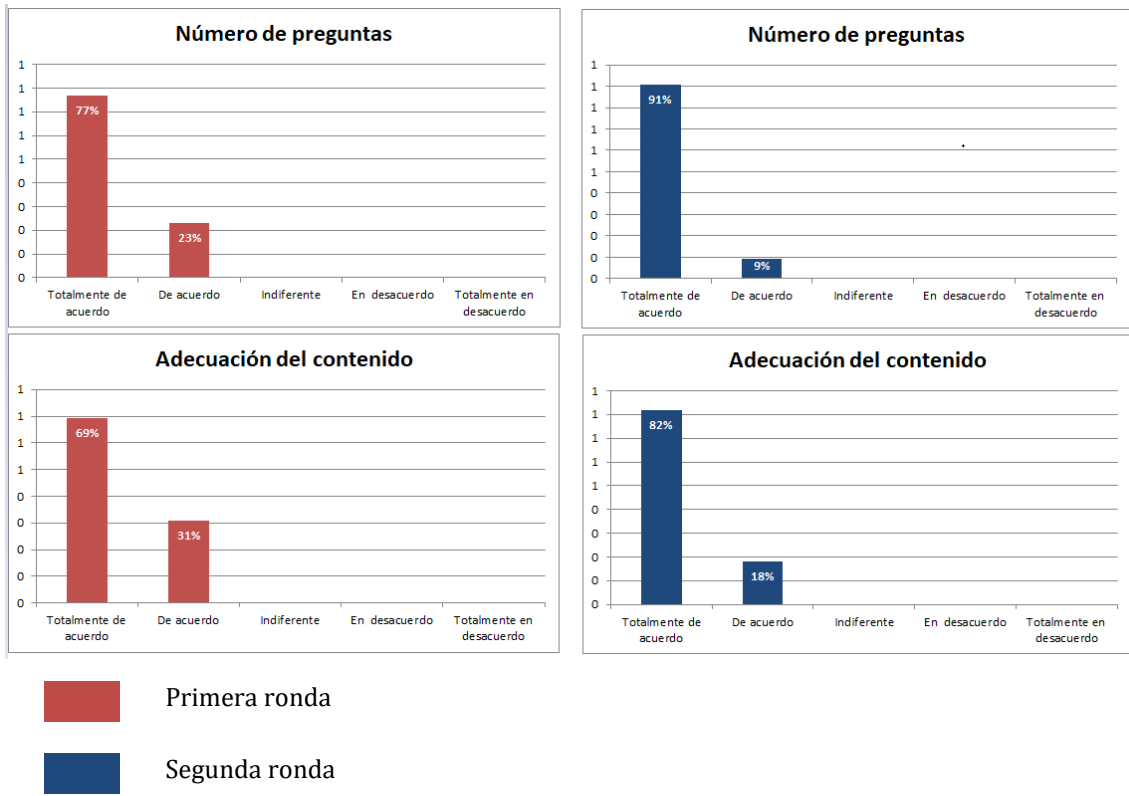


Figura 29. Porcentajes de respuesta para cada uno de los ítems y las características generales

Tabla 24. Grado de acuerdo para cada uno de los ítems en las rondas 1 y 2.

		Bloque 1							Bloque 2		
		Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9	Ítem 10
Ronda 1	Mínimo	4,00	4,00	4,00	4,00	4,00	2,00	2,00	2,00	3,00	3,00
	Máximo	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
	Cuartil 1	4,00	4,75	5,00	5,00	5,00	4,00	3,75	3,00	3,00	3,75
	Mediana	5,00	5,00	5,00	5,00	5,00	5,00	4,50	4,00	4,00	4,00
	Cuartil 3	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00
	RIR	20,00	5,00	0,00	0,00	0,00	20,00	27,78	25,00	50,00	31,25
	SD	0,51	0,48	0,44	0,44	0,38	0,97	0,99	0,93	0,86	0,80
	Media	4,62	4,69	4,77	4,77	4,85	4,46	4,15	3,77	4,08	4,15
	CV	10,97	10,24	9,19	9,19	7,75	21,68	23,76	24,59	21,15	19,27
		Bloque 1						Bloque 2			
		Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9	Ítem 10
Ronda 2	Mínimo	5,00	4,00	5,00	4,00	4,00	3,00	1,00	3,00	4,00	4,00
	Máximo	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
	Cuartil 1	5,00	4,00	5,00	5,00	5,00	4,00	4,00	4,00	5,00	4,00
	Mediana	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00
	Cuartil 3	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
	RIR	0,00	20,00	0,00	0,00	0,00	20,00	20,00	25,00	0,00	20,00
	SD	0,00	0,50	0,00	0,30	0,30	0,69	1,55	0,67	0,40	0,50
	Media	5,00	4,64	5,00	4,91	4,91	4,55	4,00	4,36	4,82	4,64
	CV	0,00	10,88	0,00	6,14	6,14	15,13	38,73	15,45	8,40	10,88

RIR: Rango Intercuartil relativo, SD: desviación estándar, CV: coeficiente de variación

Tabla 25. Grado de acuerdo para cada una de las características generales en las rondas 1 y 2

		Representatividad de una semana habitual	Claridad del contenido introductorio	Facilidad de comprensión	Claridad y sencillez del formato	División en dos bloques	Número de preguntas	Adecuación del contenido	
		Ronda 1		Mínimo	4,00	4,00	4,00	4,00	4,00
		Máximo	5,00	5,00	5,00	5,00	5,00	5,00	
		Cuartil 1	5,00	4,75	4,00	4,00	4,75	5,00	4,75
		Mediana	5,00	5,00	5,00	4,50	5,00	5,00	5,00
		Cuartil 3	5,00	5,00	5,00	5,00	5,00	5,00	5,00
		RIR	0,00	5,00	20,00	22,22	5,00	0,00	5,00
		SD	0,44	0,48	0,51	0,52	0,44	0,44	0,50
		Media	4,77	4,69	4,62	4,46	4,77	4,77	4,70
		CV	9,19	10,24	10,97	11,63	9,19	9,19	10,24
		Representatividad de una semana habitual	Claridad del contenido introductorio	Facilidad de comprensión	Claridad y sencillez del formato	División en dos bloques	Número de preguntas	Adecuación del contenido	
		Ronda 2		Mínimo	4,00	4,00	4,00	4,00	4,00
		Máximo	5,00	5,00	5,00	5,00	5,00	5,00	5,00
		Cuartil 1	5,00	5,00	5,00	4,00	5,00	5,00	5,00
		Mediana	5,00	5,00	5,00	5,00	5,00	5,00	5,00
		Cuartil 3	5,00	5,00	5,00	5,00	5,00	5,00	5,00
		RIR	0,00	0,00	0,00	20,00	0,00	0,00	0,00
		SD	0,30	0,40	0,30	0,50	0,30	0,30	0,40
		Media	4,91	4,82	4,91	4,64	4,91	4,91	4,82
		CV	6,14	8,40	6,14	10,88	6,14	6,14	8,40

RIR: Rango Intercuartil relativo, SD: desviación estándar, CV: coeficiente de variación

Una vez obtenida la versión final consensuada por los expertos, el cuestionario se sometió a un proceso de entrevistas cognitivas en un total de 5 familias que fueron seleccionadas mediante un procedimiento de muestreo estratégico no aleatorio cuyo objetivo fue maximizar la variación de los entrevistados con respecto al sexo (3 niños, 2 niñas, 5 madres, 2 padres) la edad (niños/as = 5,8(0,84) años; padres = 36(5) años), el nivel de estudios de los padres según la clasificación del INE (333) (1 enseñanza general secundaria 1º ciclo, 1 enseñanza general secundaria 2º ciclo, 1 enseñanza profesional superior, 4 estudios universitarios o equivalentes) y el entorno demográfico (2 en municipio entre 5.000 - 10.000 habitantes, 1 en municipio entre 10.000 - 20.000 habitantes y 4 en municipio de más de 100.000 habitantes). Aunque solo se seleccionaron 5 familias, cabe destacar que en dos de los casos se entrevistó a los dos progenitores de forma separada, por lo que se puede considerar que las entrevistas se realizaron en un total de 7 adultos.

El investigador a cargo del proceso (MA) estaba familiarizado con el uso de cuestionarios y la validación de instrumentos autoinformados en el ámbito de pediatría (334,335). Además, se formó específicamente en la construcción de cuestionarios de AF dirigidos a la población general infantil.

Las entrevistas duraron entre 30 y 40 minutos, teniendo en cuenta el tiempo dedicado por cada participante para responder el cuestionario y la duración de la propia entrevista. Las propuestas de mejora que surgieron a raíz de las entrevistas cognitivas fueron poco relevantes y hacían referencia principalmente a aclaraciones en el enunciado de las preguntas. Por ejemplo, una familia informó que no quedaba claro en qué ítem se contemplan los videojuegos activos (ítem 2) y que sería recomendable puntualizarlo para no incluir esta actividad dentro del CS con dispositivos electrónicos (ítem 6). Otra familia comentó que algunas escuelas implementan iniciativas para fomentar la AF en el aula con actividades tipo psicomotricidad o yoga y que

sería interesante aportar estos ejemplos dentro del enunciado del ítem 7. Por último, dos familias opinaron que sería recomendable hacer una lectura previa de los ítems antes de responder el cuestionario para no precipitarse e informar en una pregunta lo que se pide en otra posterior. Por ejemplo, para no tener en cuenta el tiempo de recreo (ítem 9) como parte de juego activo en exteriores (ítem 3). Todas estas propuestas se incluyeron en la redacción de los ítems del cuestionario para aclarar su contenido.

No se utilizó ningún enfoque específico para evaluar la información recogida en las entrevistas porque las aportaciones de las familias eran muy sencillas de analizar y no revelaron la necesidad de introducir cambios importantes en el cuestionario. Además, al tratarse de modificaciones no sustanciales se consideró que no faltaban aspectos importantes en el constructo y se dio por finalizado el proceso sin necesidad de entrevistar a más familias ni emitir otra ronda de consulta a los expertos. La versión definitiva del PAQ-YC se muestra en la Figura 30.

Los resultados de este estudio han sido publicados en la revista *Healthcare* como parte del número especial *Current Health Challenges for Child and Adolescent* (<https://doi.org/10.3390/healthcare9060655>).

Cuestionario de Actividad Física para niños/as entre 5 y 7 años

Fecha:

Nombre y apellidos del niño/a:

Edad del niño/a: años meses

Sexo del niño/a: Masculino Femenino

Curso escolar:

Parentesco con el niño/a: Padre Madre Tutor/a

Queremos saber cuánta actividad física ha hecho su hijo/a en los últimos 7 días, con qué frecuencia y durante cuánto tiempo. Esto incluye todas las actividades como deportes, gimnasia, danza... que le hacen sudar o sentirse cansado, o juegos que hacen que se acelere su respiración como saltar a la comba, correr, trepar y otros.

El cuestionario está formado por dos bloques:

- El bloque 1 contiene preguntas que hacen referencia a las actividades que su hijo/a realiza en su tiempo libre y durante los desplazamientos. Estas preguntas podrá responderlas usted solo/a.
- El bloque 2 contiene preguntas relacionadas con la actividad física que su hijo/a hace en la escuela. Para responder a las preguntas del bloque 2, le pedimos que se sienta con su hijo/a y que le pregunte para marcar la respuesta. Si no estuviera seguro de la respuesta, le rogamos que consulte la información con los profesores de la escuela.

Le agradecemos total sinceridad en sus respuestas. La información que nos proporcione será tratada confidencialmente y de forma anónima.

Recuerde:

- Marque con una **X** la casilla que se corresponda con su respuesta
- Debe responder a todas las preguntas. No hay preguntas correctas o incorrectas

PREGUNTAS DEL BLOQUE 1

Actividad física en el tiempo libre

1. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha hecho actividades dirigidas por monitor/entrenador en horario extraescolar que requieran estar activo (por ejemplo, practicando deportes, danza...)? No tenga en cuenta en esta pregunta el tiempo en la escuela, como el recreo o la Educación Física.

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a ha hecho más de una actividad extraescolar el mismo día, sume el tiempo que ha dedicado a cada actividad para obtener el tiempo total. *Por ejemplo minutos: el lunes ha ido 1 hora a natación y 45 minutos a patinaje. El tiempo total es: 1 hora + 45 minutos = 1 hora y 45*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

2. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha jugado activamente en casa (por ejemplo, bailando, corriendo, empujando juguetes, jugando a videojuegos activos como la Nintendo Wii...) o en instalaciones de juego interiores (por ejemplo, en una ludoteca, en una piscina de bolas...)?

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a ha hecho más de una actividad el mismo día, sume el tiempo que ha dedicado a cada actividad para obtener el tiempo total. *Por ejemplo: el lunes ha jugado en la ludoteca 30 minutos y en casa ha bailando mientras escuchaba música durante 20 minutos. El tiempo total es 30 minutos + 20 minutos = 50 minutos.*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

3. En los últimos 7 días, ¿qué días y durante cuánto tiempo al día diría usted que su hijo/a ha jugado activamente en exteriores (por ejemplo, en el jardín, en el parque...)? No tenga en cuenta las actividades extraescolares dirigidas que se han incluido en la pregunta 1 ni el tiempo de recreo en la escuela.

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a ha hecho más de una actividad el mismo día, sume el tiempo que ha dedicado a cada actividad para obtener el tiempo total. *Por ejemplo: el lunes ha jugado en el parque 30 minutos y en casa ha jugado en el jardín durante 20 minutos. El tiempo total es 30 minutos + 20 minutos = 50 minutos.*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

Transporte activo

4. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a se ha desplazado de un sitio a otro de forma activa (por ejemplo, caminando, en bicicleta, en patinete...) sin utilizar ningún medio de transporte (ni coche, ni autobús, ni metro...)? Incluya en esta pregunta si el niño/a va caminando a la escuela y otros posibles desplazamientos (por ejemplo, ir de compras, ir a las actividades extraescolares...).

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a se ha desplazado varias veces al día sin utilizar ningún medio de transporte, sume el tiempo que ha tardado en desplazarse cada una de las veces para obtener el tiempo total. *Por ejemplo: el lunes ha ido y vuelto del colegio caminando (10 minutos de trayecto de ida y 10 minutos de trayecto de vuelta) y después ha ido y vuelto al parque en bicicleta (15 minutos de trayecto de ida y 15 minutos de trayecto de vuelta). El tiempo total es 10 minutos + 10 minutos + 15 minutos + 15 minutos = 50 minutos.*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

Comportamiento sedentario en el tiempo libre

5. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha estado sentado jugando con juguetes o sin moverse haciendo actividades como colorear, dibujar, hacer manualidades, hacer los deberes...? No tenga en cuenta las horas en la escuela ni el tiempo con dispositivos electrónicos (por ejemplo, la TV, el ordenador, los videojuegos...).

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a ha hecho más de una actividad el mismo día, sume el tiempo que ha dedicado a cada actividad para obtener el tiempo total. *Por ejemplo: el lunes ha coloreado durante 15 minutos y ha hecho los deberes durante 1 hora. El tiempo total es 15 minutos + 1 hora = 1 hora y 15 minutos.*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

6. En los últimos 7 días, ¿qué días y durante cuánto tiempo diría usted que su hijo/a ha estado sentado o sin moverse haciendo actividades como ver la TV, hacer los deberes con ordenador o Tablet, jugar con un videojuego, jugar al ordenador, móvil o Tablet? No tenga en cuenta las horas en la escuela ni los videojuegos activos como la Nintendo Wii.

Marque con una **X** la casilla que se corresponda con su respuesta o escriba el número de horas y minutos en caso de que sean más de 2 horas.

Si su hijo/a ha hecho más de una actividad el mismo día, sume el tiempo que ha dedicado a cada actividad para obtener el tiempo total. *Por ejemplo: el lunes ha visto la TV durante 1 hora y ha jugado con un videojuego durante 45 minutos. El tiempo total es 1 hora + 45 minutos = 1 hora y 45 minutos.*

	Nada o menos de 10 minutos	Entre 11 y 30 minutos	Entre 31 y 59 minutos	Entre 1 hora y 1 hora y 29 minutos	Entre 1 hora y media y 2 horas	Más de 2 horas (escribir cuánto)
Lunes					 horas y minutos
Martes					 horas y minutos
Miércoles					 horas y minutos
Jueves					 horas y minutos
Viernes					 horas y minutos
Sábado					 horas y minutos
Domingo					 horas y minutos

PREGUNTAS DEL BLOQUE 2

Recuerde que para responder a estas preguntas debe preguntarle a su hijo/a. Si no está seguro de las respuestas, debe consultar la información con los profesores de la escuela.

Actividad física en la escuela

7. ¿Cuántas horas de Educación Física o actividades similares (por ejemplo, psicomotricidad, yoga, natación...) hacen a la semana en el curso de su hijo/a?

En el curso de mi hijo/a hacen horas y minutos de Educación Física a la semana.

8. ¿Su hijo/a realiza habitualmente las clases de Educación Física en la escuela?

- Mi hijo/a no hace Educación Física
- Mi hijo/a solo hace Educación Física cuando hay que hacer poco esfuerzo físico*
- Mi hijo/a siempre hace Educación Física

* *Esfuerzo físico incluye todas las actividades que le hacen sudar y sentirse cansado o que hacen que se acelere su respiración como correr, saltar, hacer lanzamientos, practicar algún deporte...*

9. ¿Qué hace habitualmente su hijo/a durante el recreo en la escuela?

- Estar sentado (hablar, leer, jugar a juegos sin moverse, deberes...)
- Correr y jugar a juegos que impliquen poco esfuerzo físico*
- Correr y jugar intensamente todo el tiempo

Pausa a mediodía

10. ¿Qué hace habitualmente su hijo/a durante la pausa a mediodía (además de comer)? Responda a esta pregunta tanto si el niño se queda a comer en el comedor escolar como si va a comer a casa.

- Estar sentado (hablar, leer, jugar a juegos sin moverse, ver la TV, hacer deberes...)
- Correr y jugar a juegos que impliquen poco esfuerzo físico*
- Correr y jugar intensamente todo el tiempo

¿Estuvo enfermo su hijo/a la última semana o algo impidió que hiciera las actividades habituales?

- Sí
- No

Figura 30. Versión final del PAQ-YC

4.6.1.1. Prueba de campo

La prueba de campo contó con un total de 115 participantes entre 5 y 7 años que respondieron al cuestionario en una semana típica del curso escolar. Se excluyeron del análisis 11 respuestas que reflejaban que el niño/a había estado enfermo durante la última semana o algún incidente había impedido que hiciera las actividades habituales. De las 104 observaciones válidas, 47 (45,19%) eran niños y 57 (54,81%) eran niñas. La edad media fue de 75,51 (0,96) meses. La distribución por cursos escolares fue de 27 (26%) participantes en P5, 45 (43,2%) en 1º de Educación Primaria y 32 (30,8%) en 2º de Educación Primaria. En relación al parentesco del respondedor, en 20 (19,23%) de los casos fueron los padres y en 84 (80,77%) fueron las madres.

En la Tabla 26 se muestra la distribución de los resultados del cuestionario para cada uno de los ítems que contribuyen a la puntuación, para las puntuaciones del bloque 1 y 2, para la puntuación total y para los porcentajes de AF y CS discrecional. En cada caso, se presenta el rango de puntuación, la media, la desviación estándar, la mediana, el rango intercuartílico, el mínimo y el máximo alcanzados durante la prueba.

Ninguno de los ítems del PAQ-YC mostró un impacto despreciable en la puntuación total del cuestionario. Excepto para el ítem 4, todas las preguntas obtuvieron representación de cada uno de los valores de puntuación posibles. En relación a la presencia de valores extremos en la puntuación, se consideró que no existía ni efecto techo ni efecto suelo ya que ninguno de los participantes mostraba la máxima o mínima puntuación.

La prueba de campo confirmó los resultados de las entrevistas con expertos y con muestras representativas de la población diana en relación a la relevancia de los ítems que componen el PAQ-YC, por lo que se decidió no hacer ninguna modificación a la versión final.

Tabla 26. Distribución de las puntuaciones del PAQ-YC en la prueba de campo

Variable	Rango	Mín-Máx	Media (SD)	Mediana (RIQ)
Ítem 1	0 - 5	0 - 5	1,68 (1,22)	2 (1,00)
Ítem 2	0 - 5	0 - 5	1,34 (1,01)	1 (1,00)
Ítem 3	0 - 5	0 - 5	3,13 (1,35)	3 (2,00)
Ítem 4	0 - 5	0 - 4	1,16 (0,76)	1 (1,00)
Ítem 8*	0 - 3	0 - 3	1,73 (0,53)	2 (0,75)
Ítem 9	0 - 2	0 - 2	1,56 (0,54)	2 (1,00)
Ítem 10	0 - 2	0 - 2	0,90 (0,82)	1 (2,00)
Bloque 1	0 - 20	1 - 14	7,32 (2,41)	7 (3,75)
Bloque 2	0 - 7	1 - 7	4,19 (1,24)	4 (2,00)
Total	0 - 27	5- 17	11,51 (2,67)	11 (3,00)
% AF discrecional	0 - 100	17,32 - 80,62	49,91 (12,63)	50,22 (18,53)
% CS discrecional	0 - 100	19,38 - 82,68	50,09 (12,63)	49,78 (18,53)

Mín: mínimo, Máx: máximo; SD: desviación estándar; RIQ: rango intercuartílico, AF: actividad física, CS: comportamiento sedentario, * con factor de corrección del ítem 7

4.6.2. Fiabilidad y error de medición

Los estudios de fiabilidad y error de medición se realizaron con las respuestas de un total de 83 participantes, de los cuales 39 (46,99%) eran niños y 44 (53,01%) eran niñas. La edad media fue de 75,13 (1,01) meses. La distribución por cursos escolares fue de 21 (25,30%) participantes en P5, 40 (48,20%) en 1º de Educación Primaria y 22 (26,50) en 2º de Educación Primaria. En 18 (21,69%) de los casos el respondedor fue el padre y en los 65 (78,31%) restantes fueron las madres.

En la Tabla 27 se muestra la distribución de los resultados del cuestionario en el test y retest para cada uno de los ítems que contribuyen a la puntuación, para las puntuaciones del bloque 1 y 2, para la puntuación total y para los porcentajes de AF y CS discrecional. En cada caso, se presenta el rango de

puntuación, la media, la desviación estándar, la mediana, el rango intercuartílico, el mínimo y el máximo alcanzados durante la prueba.

Tabla 27. Distribución de las puntuaciones del PAQ-YC en el estudio de fiabilidad y error de medición

	Rango	Observación	Min-Max	Media (SD)	Mediana (RIQ)
Ítem 1	0 - 5	Test	0 - 5	1,73 (1,19)	2 (1,00)
		Retest	0 - 5	1,72 (1,20)	2 (1,00)
Ítem 2	0 - 5	Test	0 - 5	1,37 (1,03)	1 (1,00)
		Retest	0 - 4	1,36 (0,97)	1 (1,00)
Ítem 3	0 - 5	Test	0 - 5	3,08 (1,38)	3 (2,00)
		Retest	0 - 5	3,08 (1,36)	3 (2,00)
Ítem 4	0 - 5	Test	0 - 4	1,22 (0,78)	1 (1,00)
		Retest	0 - 3	1,22 (0,78)	1 (1,00)
Ítem 8*	0 - 3	Test	0 - 3	1,69 (0,56)	2 (1,00)
		Retest	0 - 3	1,69 (0,56)	2 (1,00)
Ítem 9	0 - 2	Test	0 - 2	1,55 (0,52)	2 (1,00)
		Retest	0 - 2	1,53 (0,55)	2 (1,00)
Ítem 10	0 - 2	Test	0 - 2	0,89 (0,81)	1 (2,00)
		Retest	0 - 2	0,90 (0,82)	1 (2,00)
Bloque 1	0 - 20	Test	1 - 14	7,41 (2,51)	8 (4,00)
		Retest	1 - 14	7,39 (2,46)	7 (4,00)
Bloque 2	0 - 7	Test	1 - 7	4,13 (1,26)	4 (2,00)
		Retest	1 - 7	4,12 (1,31)	4 (2,00)
Total 1	0 - 27	Test	5 - 17	11,54 (2,82)	11 (4,00)
		Retest	5 - 17	11,51 (2,81)	11 (4,00)
% AF discrecional	0 - 100	Test	17,32 - 80,62	50,23 (13,12)	51,24 (18,89)
		Retest	18,03 - 79,59	49,60 (13,27)	50,54 (17,35)
% CS discrecional	0 - 100	Test	19,38 - 82,68	49,77 (13,12)	48,76 (18,89)
		Retest	20,41 - 81,97	50,40 (13,27)	49,46 (17,35)

Mín: mínimo, Máx: máximo, SD: desviación estándar, RIQ: rango intercuartílico, AF: actividad física, CS: comportamiento sedentario, * con factor de corrección del ítem 7

Los resultados de la fiabilidad y el error de medición entre la primera y la segunda administración del cuestionario se muestran en la Tabla 28 y en la Tabla 29. Para los ítems que contribuyen a la puntuación del PAQ-YC se obtuvo un índice Kappa que oscila entre 0,855 y 1,000. El error de medición expresado como el SEM mostró resultados que varían entre 0,00 y 0,05. Así mismo, se observó un PoA superior al 89% en todas las variables analizadas, siendo del 100% para el ítem 8 corregido (Tabla 28). Para los resultados del Bloque 1, del Bloque 2, de la puntuación total, del porcentaje de AF y del porcentaje de CS discrecional, se obtuvieron ICC superiores a 0,9. Por ello, en todos los casos se consideró que la fiabilidad era excelente en base a los criterios de Fleiss et al. (330). El error de medición expresado como el SEM presentó valores de entre 0,28 y 0,61 para la puntuación de los bloques 1 y 2 y para los resultados de la puntuación total, mientras que para los porcentajes de AF y CS discrecional el SEM fue de 2,88. El PoA entre la primera y la segunda administración fue del 79,52%, del 93,98% y del 75,90% para el bloque 1, el bloque 2 y la puntuación total, respectivamente (Tabla 29).

Tabla 28. Fiabilidad y error de medición de los ítems que contribuyen a la puntuación del PAQ-YC

	Kappa	SEM	PoA
Ítem 1	0,855	0,04	89,16%
Ítem 2	0,880	0,04	92,77%
Ítem 3	0,908	0,04	92,77%
Ítem 4	0,883	0,05	93,98%
Ítem 8*	1,000	0,00	100,00%
Ítem 9	0,905	0,05	95,18%
Ítem 10	0,945	0,03	96,39%

SEM: error estándar de medición, PoA: porcentaje de acuerdo, * con factor de corrección del ítem 7

Tabla 29. Fiabilidad y error de medición de la puntuación del Bloque 1, Bloque 2, Total, porcentaje de actividad física y porcentaje de comportamiento sedentario discrecional

	Rango	ICC	IC 95%	SEM	PoA
Bloque 1	0 - 20	0,983	[0,974 - 0,989]	0,54	79,52%
Bloque 2	0 - 7	0,991	[0,986 - 0,994]	0,28	93,98%
Total	0 - 27	0,986	[0,979 - 0,991]	0,61	75,90%
% AF discrecional	0 - 100	0,990	[0,984 - 0,993]	2,88	
% CS discrecional	0 - 100	0,990	[0,984 - 0,993]	2,88	

ICC: índice de concordancia intraobservador, IC 95%: intervalo de confianza al 95%, SEM: error estándar de medición, PoA: porcentaje de acuerdo, AF: actividad física, CS: comportamiento sedentario, * con factor de corrección del ítem 7

4.6.3. Validez de constructo

En el estudio de validez de constructo para el PAQ-YC participaron un total de 42 voluntarios. Se excluyeron del análisis los datos de 6 participantes, de los cuales 3 fueron confinados por sospecha de Covid-19 durante el período de medición, 2 no cumplían el criterio de días válidos y 1 no respondió el cuestionario tras la semana de medición del acelerómetro. Por tanto, el análisis se realizó con los datos de 36 observaciones válidas, de las cuales 21 (58,33%) eran niños y 15 (41,66%) eran niñas. La edad media fue de 79,94 (1,01) meses. La distribución por cursos escolares fue de 4 (11,11%) participantes en P5, 21 (58,33%) en 1º de Educación Primaria y 11 (30,56%) en 2º de Educación Primaria. En 8 (22,22%) de los casos el respondedor fue el padre y en los 28 (77,78%) restantes fueron las madres. Con respecto al cumplimiento en el uso del acelerómetro, 21 (58,33%) participantes presentaron datos válidos para los 7 días de medición, mientras que para 12 (33,33%) participantes se excluyeron los resultados de un día entre semana y en 3 (8,33%) casos se retiraron los datos de un día del fin de semana. La media de minutos de uso del dispositivo fue de 828 (21,88) minutos, con un

cumplimiento ligeramente superior para los días entre semana (835 (22,29) minutos) en comparación con el fin de semana (810 (1,02) minutos).

En la Tabla 30 se muestra la distribución del cuestionario para cada uno de los ítems que contribuyen a la puntuación, para las puntuaciones del bloque 1 y 2, para la puntuación total y para los porcentajes de AF y CS discrecional de los participantes del estudio de validez de constructo. En cada caso, se presenta el rango de puntuación, la media, la desviación estándar, la mediana, el rango intercuartílico, el mínimo y el máximo alcanzados durante la prueba. Así mismo, para los resultados de la puntuación total y los porcentajes de AF y CS discrecional se muestra el p valor del test de Shapiro-Wilk que indica que los datos siguen una distribución normal.

Tabla 30. Distribución de las puntuaciones del PAQ-YC en el estudio de validez de constructo

Variable	Rango	Mín-Máx	Media (SD)	Mediana (RIQ)	p valor (Shapiro-Wilk)
ítem 1	0 - 5	0 - 4	1,25 (1,23)	1 (2,00)	
ítem 2	0 - 5	0 - 2	0,94 (0,79)	1 (2,00)	
ítem 3	0 - 5	0 - 5	3,25 (1,38)	3 (2,75)	
ítem 4	0 - 5	0 - 2	1,03 (0,65)	1 (0,00)	
ítem 8*	0 - 3	1 - 2	1,97 (0,17)	2 (0,00)	
ítem 9	0 - 2	1 - 2	1,67 (0,48)	2 (1,00)	
ítem 10	0 - 2	0 - 2	1,08 (0,84)	1 (2,00)	
Puntuación bloque 1	0 - 20	1 - 11	6,47 (2,20)	6 (3,00)	
Puntuación bloque 2	0 - 7	3 - 6	4,58 (1,02)	4 (1,75)	
Puntuación total	0 - 27	5 - 16	11,06 (2,44)	11 (3,75)	
% AF discrecional	0 - 100	20,10 - 63,93	43,81 (10,04)	44,56 (13,19)	0,470
% ST discrecional	0 - 100	36,07 - 79,90	56,19 (10,04)	55,45 (13,19)	0,470

Mín: mínimo, Máx: máximo, SD: desviación estándar, RIQ: rango intercuartílico, AF: actividad física, CS: comportamiento sedentario, * con factor de corrección del ítem 7

En la Tabla 31 se muestra el análisis descriptivo de los registros totales, durante el tiempo discrecional y durante las horas en la escuela de los dispositivos Actigraph wGT3x-BT (ActiGraph Corp., Pensacola, FL). Para cada variable, se presenta la media, la desviación estándar, la mediana, el rango intercuartílico, el mínimo y el máximo alcanzados durante la prueba. Así mismo, para los recuentos por minuto (cpm) totales y los porcentajes de AF y CS discrecional se muestra el p valor del test de Shapiro-Wilk que indica que los datos siguen una distribución normal.

Tabla 31. Distribución de los registros de los dispositivos Actigraph wGT3x-BT (ActiGraph Corp., Pensacola, FL)

	Variable	Mín-Máx	Media (SD)	Mediana (RIQ)	p valor (Shapiro-Wilk)
Total	cpm	848,10 - 1.923,30	1.326,24 (291,79)	1.344,70 (498,70)	0,215
	% CS	0,45 - 0,72	0,57 (0,08)	0,56 (0,16)	
	% AF ligera	0,22 - 0,43	0,33 (0,06)	0,34 (0,10)	
	% AF moderada	0,03 - 0,10	0,06 (0,02)	0,06 (0,03)	
	% AF vigorosa	0,01 - 0,07	0,03 (0,01)	0,03 (0,01)	
	% AF total	0,28 - 0,55	0,43 (0,08)	0,44 (0,16)	
Discrecional	cpm	747,60 - 2.264,50	1.391,68 (350,44)	1.323,85 (506,93)	0,364
	% CS	38,76 - 72,81	56,63 (8,69)	57,45 (15,74)	
	% AF ligera	0,23 - 0,43	0,33 (0,06)	0,33 (0,10)	
	% AF moderada	0,03 - 0,13	0,07 (0,02)	0,06 (0,03)	
	% AF vigorosa	0,01 - 0,09	0,04 (0,02)	0,03 (0,02)	
	% AF total	27,19 - 61,23	43,37 (8,69)	42,56 (15,73)	
Escuela	cpm	896,00 - 1.986,50	1.408,56 (271,69)	1.419,50 (406,28)	0,365
	% CS	0,37 - 0,67	0,54 (0,07)	0,54 (0,12)	
	% AF ligera	0,27 - 0,46	0,37 (0,05)	0,36 (0,08)	
	% AF moderada	0,03 - 0,11	0,06 (0,02)	0,06 (0,02)	
	% AF vigorosa	0,01 - 0,06	0,03 (0,01)	0,03 (0,02)	
	% AF total	0,33 - 0,63	0,46 (0,07)	0,46 (0,12)	

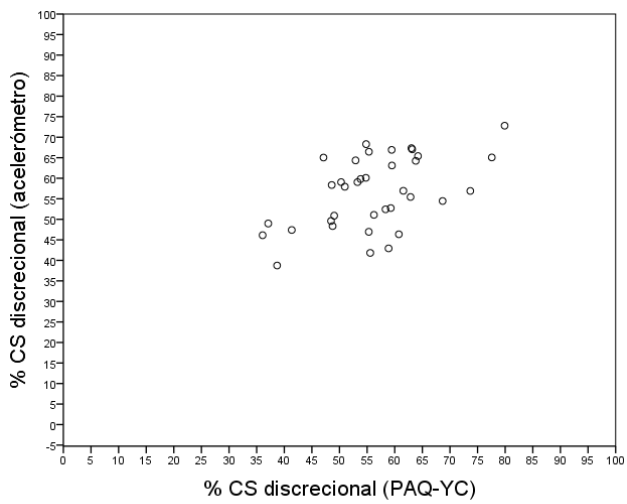
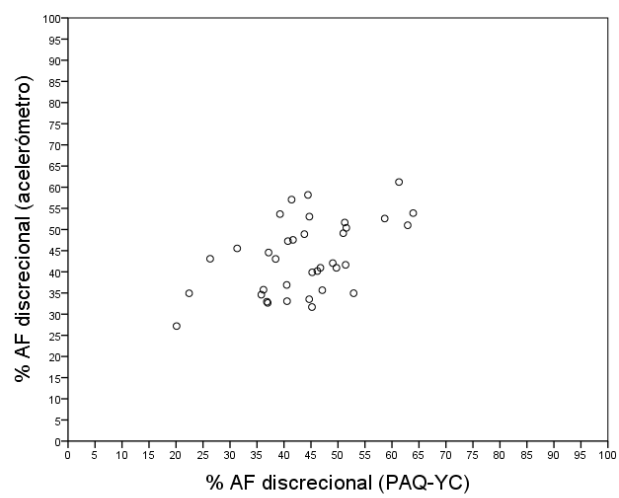
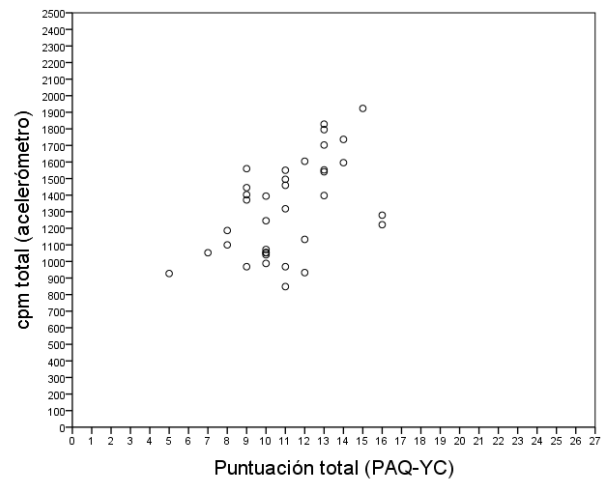
Mín: mínimo, Máx: máximo, SD: desviación estándar, RIQ: rango intercuartílico, AF: actividad física, CS: comportamiento sedentario, cpm: recuentos por minuto

Los resultados de los análisis para la validez de constructo se presentan en la Tabla 32. Se calculó el coeficiente de correlación de Pearson para las variables resultado del cuestionario (puntuación total, porcentaje de AF y CS discrecional) y sus equivalentes en los registros del acelerómetro. En todos los casos, se consideró que existía una validez de constructo moderada que cumple los criterios de suficiencia propuestos por Terwee et al. (192) pero no alcanza el nivel mínimo exigible de Hidding et al. (13) de $r = 0,6$ para proporcionar una evidencia positiva de nivel 1. Así mismo, en la Figura 31 se presentan los gráficos de dispersión entre las variables analizadas.

Tabla 32. Validez de constructo por comparación de métodos entre el PAQ-YC y acelerometría Actigraph wGT3X-BT

Variable	Coeficiente de correlación	p valor
cpm total (acelerómetro) Puntuación total (PAQ-YC)	0,517**	0,001
% AF discrecional (acelerómetro) % AF discrecional (PAQ-YC)	0,513**	0,001
% CS discrecional (acelerómetro) % CS discrecional (PAQ-YC)	0,511**	0,001

cpm: recuentos por minuto, PAQ-YC: Physical Activity Questionnaire – Young Children, AF: actividad física, CS: comportamiento sedentario, ** correlación significativa



cpm: recuentos por minuto, PAQ-YC: Physical Activity Questionnaire – Young Children, AF: actividad física, CS: comportamiento sedentario

Figura 31. Gráficos de dispersión entre las variables analizadas para la validez de constructo del PAQ-YC

4.7. Discusión

4.7.1. Desarrollo y validez de contenido

La investigación cualitativa para respaldar el desarrollo de un nuevo cuestionario debe tener una base científica sólida y documentable y debe realizarse con el rigor requerido para toda investigación (336). En primer lugar, cuando se publica un nuevo cuestionario, ha de quedar claramente descrito el constructo específico que se pretende recoger (por ejemplo, AF total, AF en el tiempo libre, AF de intensidad moderada-vigorosa, etc.), la población diana (por ejemplo, población general, sexo, rango de edad, condiciones específicas, etc.) y el objetivo de valoración (por ejemplo, discriminativo, evaluativo o predictivo) (201). En este sentido, el PAQ-YC es un cuestionario genérico diseñado para medir la AF total con fines discriminativos durante el período lectivo en población de 5 a 7 años.

En relación al constructo, el PAQ-YC pretende capturar la totalidad de la AF practicada en todos los rangos de intensidad y en todos los dominios. En este contexto, de los cuestionarios disponibles para la franja de edad de interés en este estudio, destaca la gran variedad de constructos que se miden. Por ejemplo, el NPAQ fue diseñado para evaluar las preferencias de actividades bajo el supuesto de que se asocian con los niveles de AF (215,337), mientras que el cuestionario de AF para padres y profesores de Harro (210,303), el ACTIVITY (254) y el cuestionario de AF del SAYCARE (297) cuantifican la AF total practicada en todos los dominios y en todos los niveles de intensidad, por lo que en este aspecto serían muy similares al PAQ-YC. Por otro lado, los cuestionarios que incluyen al menos dos años de la franja de edad de la población diana del PAQ-YC (5-6 años o 6-7 años) recogen constructos más acotados, centrándose fundamentalmente en la medición de la AF de mayor intensidad. Por ejemplo, el CLASS (205), el cuestionario *School Health Action, Planning and Evaluation System* (SHAPES) (338), el cuestionario de AF y CS para padres de Bringolf-Isler (339) o el CHMS (322) informan sobre el tiempo

dedicado a la AFMV, mientras que la Encuesta de Salud para el Cuestionario de AF de Inglaterra mide la AFMV, pero no registra las actividades del dominio escuela (340). En comparación con los cuestionarios del grupo IPAQ para la franja pediátrica, tanto el EY-PAQ (224), como el PAQ-C (208,253) y el PAQ-A (341) miden la AF en el contexto de intensidades de moderadas a vigorosas, a diferencia del PAQ-YC que recoge la AF en todos los rangos de intensidad.

En cuanto a la población a la que se dirige, las guías sobre construcción de instrumentos indirectos resaltan la importancia de definir para quién se desarrolló el cuestionario como criterio indispensable para determinar la relevancia y exhaustividad del contenido y estudiar la posible extrapolación del cuestionario a otras aplicaciones (172,177,193). Los expertos recomiendan que se proporcione una clara descripción con respecto al tipo y características importantes de la patología (por ejemplo, estadio, comorbilidades, etc.), las características demográficas (por ejemplo, grupo de edad, sexo) u otros aspectos relevantes (243). En el caso del PAQ-YC, al tratarse de un cuestionario genérico, únicamente se definieron las características demográficas de la población, es decir, niños/as entre 5 y 7 años.

En este punto, es importante señalar la dificultad de diseñar cuestionarios específicos para esta población, ya que incluye un rango de edad a caballo entre la Educación Infantil y la Educación Primaria que combina momentos en los que los niños no están a cargo directo de los padres, pero tampoco disponen de los suficientes recursos como para responder al cuestionario de forma autónoma. Los expertos señalan que el funcionamiento cognitivo más bajo de los niños en comparación con los adultos reduce su capacidad para interpretar correctamente las preguntas y recordar con precisión la actividad en términos de intensidad, frecuencia y, especialmente, duración (15,292). Por esta razón, los cuestionarios de AF dirigidos a niños deben diseñarse de tal manera que el impacto de las habilidades cognitivas, de memoria y de estimación se reduzca a un mínimo aceptable (214). Aun así, los expertos

recomiendan que el autoinforme se aplique únicamente a partir del umbral de los 10 años, debido a que las limitaciones cognitivas de los niños más pequeños comportan grandes inexactitudes (7,252,301,342,343).

Esta casuística se puede solventar de diferentes formas. Lo más habitual es que se generen informes indirectos por parte de los cuidadores, que pueden ser los padres y/o los profesores. Aunque algunos autores sugieren que este enfoque puede añadir una nueva fuente de error, en general, se considera que reduce significativamente los sesgos de memoria causados por las limitaciones cognitivas de los niños (7,172,300,301). De los cuestionarios que se pueden administrar al rango de edad de 5 a 7 años, el NPAQ (215,337), el cuestionario de AF para padres y profesores de Harro (210,303) o el cuestionario del SAYCARE (297) son ejemplos de reportes por parte de los cuidadores. Así mismo, el CLASS (5-6 años) (205), el cuestionario para padres de Bringolf-Isler (6-10 años) (339), la Encuesta de Salud para el Cuestionario de AF de Inglaterra (6-7 años) (340) o el CHMS (6-11 años) (322) son también informes indirectos. Otra posibilidad es crear dos versiones del instrumento, una para ser respondida por los niños y otra por los padres. Aunque este enfoque no es habitual en la medición de la AF, es común en otras medidas en las que se valora tanto la percepción individual como la percepción de los cuidadores. Por ejemplo, la escala Pediatric Quality of Life Inventory (PedsQL) dispone de dos versiones para evaluar la calidad de vida autopercibida por niños de 5 a 7 años y sus padres (344,345). Por último, y aunque no está recomendado, hay algún cuestionario de AF que adapta las preguntas para que sean respondidas por los propios niños sin ayuda de los adultos responsables. Por ejemplo, el ACTIVITY (dirigido a menores de 10 años) (254) y el cuestionario SHAPES (6-12 años) (338) miden la AF a través del método del autoinforme. Con respecto a los cuestionarios del grupo IPAQ para la franja pediátrica, el EY-PAQ (224) es informado por los padres mientras el PAQ-C (208,253) y el PAQ-A (341) son instrumentos de autoinforme.

En el caso del PAQ-YC, se optó por diseñar un cuestionario para ser respondido por los padres en el que se pretende además compendiar la información proporcionada por el niño y/o el profesor en el dominio escuela. Por ello, al inicio del bloque 2, se aclara “*Recuerde que para responder a estas preguntas debe preguntarle a su hijo/a. Si no está seguro de las respuestas, debe consultar la información con los profesores de la escuela*”. Esta decisión se tomó en base a la problemática descrita acerca de la dificultad de los padres para reportar la totalidad de la AF de sus hijos, ya que se desarrolla en entornos en los que no siempre están presentes (300,301).

Por otro lado, el diseño del PAQ-YC se ajustó a su aplicación prevista con fines discriminativos, con la intención de que los resultados del cuestionario sirvan para clasificar a los individuos en categorías de AF según el cumplimiento de las recomendaciones internacionales (3,225). En su forma actual, la mayoría de los cuestionarios de AF se han desarrollado con esta finalidad, aunque habitualmente se utilizan con objetivos diferentes, especialmente con fines evaluativos pre y post intervención. Por ejemplo, el NPAQ (215,337) o el IPAQ (203) se desarrollaron como índices discriminativos, aunque a menudo se utilizan para evaluar cambios en los comportamientos de AF sin pruebas de una suficiente capacidad de respuesta (346).

En relación al período de recuerdo, las respuestas de los cuestionarios dependen del sesgo de memoria para definir los patrones de actividad (111,172,178), especialmente complejos en pediatría debido a la característica intermitente de la AF (15,56). Por tanto, los períodos de recuerdo largos imponen una gran carga cognitiva y no son adecuados para medir la AF a edades tempranas. Incluso en adultos, se ha descrito que presentan problemas para cuantificar el tiempo que se dedica a cada una de las actividades (187). En cuanto a la noción de “habitual”, hay discrepancias entre los expertos sobre su idoneidad, aunque algunos autores consideran

que el término genera confusión por la dificultad de estimar lo que es habitual (187,203).

En este contexto, el PAQ-YC se diseñó con un período de recuerdo que hace referencia a las actividades practicadas durante la última semana, entendiéndose que pueden ser representativas de la condición habitual. En los cuestionarios destinados a medir la AF en la misma o muy cercana franja de edad, los períodos de recuerdo son muy variados. Por ejemplo, el NPAQ pregunta por la AF en los 6 meses previos (215,337); el EY-PAQ recoge las actividades practicadas durante el último mes (224); el cuestionario del SAYCARE (297), la Encuesta de Salud para el Cuestionario de AF de Inglaterra (340) y el cuestionario SHAPES (338) se refieren a los últimos 7 días; el CLASS (205), el CHMS (322), el PAQ-C (208,253) y el PAQ-A (341) aluden a una semana típica; el cuestionario para padres y profesores de Harro (210,303) y el ACTIVITY (254) miden la AF del día anterior y el cuestionario para padres de Bringolf-Isler recoge dos días típicos (entre semana y fin de semana) (339).

Con respecto a la validez del PAQ-YC durante el período lectivo, es evidente que, si bien los cuestionarios de recuerdo dan como resultado información detallada sobre los patrones de AF, existe una limitación en la representatividad del comportamiento de AF durante todo el año (168). Generalmente, los cuestionarios de AF contienen preguntas relacionadas con el dominio escuela, lo que los convierte en instrumentos aplicables solo durante los meses lectivos. Este es el caso de los cuestionarios CLASS (205), ACTIVITY (254), SAYCARE (297), PAQ-C (208,253), PAQ-A (341) o el cuestionario para padres y profesores de Harro (210,303), entre otros. Esto es importante ya que si surge la necesidad de evaluar el nivel de AF durante otra temporada (por ejemplo, en verano) se ha de buscar un cuestionario específico en lugar de omitir ítems y alterar la versión original (292). Una posible solución a esta casuística sería administrar cuestionarios con

períodos de recuerdo amplios que puedan cubrir la variación estacional, pero como ya se ha comentado, no se recomiendan en edades tempranas. Por el contrario, existen cuestionarios que omiten deliberadamente las preguntas relativas al dominio escuela, como la Encuesta de Salud para el Cuestionario de AF de Inglaterra (340).

En relación a la composición de un cuestionario mono-ítem o formado por múltiples ítems, se decidió formular un cuestionario de recuerdo integrado por un total de 10 preguntas para recoger actividades más específicas y con mayor grado de detalle. Esta decisión está en consonancia con la idea de que los constructos complejos se miden mejor con instrumentos multi-ítem (111,174,201). Concretamente, se descartó diseñar un cuestionario de historia global porque se pretendía capturar el detalle de las actividades y con un número reducido de ítems no habría sido posible describir los patrones de AF (58,100,174). Así mismo, se evitó construir un cuestionario de historia cuantitativa porque requiere un coste de tiempo elevado (58,174) y habitualmente se asocia a amplios períodos de recuerdo que pueden no detectar los cambios en los patrones de AF que se dan durante el crecimiento (30,56,90). A pesar de ello, algunos cuestionarios dirigidos a esta franja de edad comportan un gran número de ítems, como el cuestionario del SAYCARE, que está formado por 47 preguntas (297), el cuestionario SHAPES, que consta de 45 ítems (338) o el cuestionario para padres y profesores de Harro, que contiene 12 preguntas para los padres y de 16 a 17 preguntas para los profesores (210). Sin embargo, la mayor parte de cuestionarios de AF comentados por el momento encajan dentro de la clasificación de cuestionarios de recuerdo y engloban de 7 a 15 preguntas (215,253,254,322,339,341).

Para facilitar el proceso de obtención de conceptos representativos del constructo, el desarrollo del PAQ-YC fue precedido de un análisis de la literatura científica que permitió definir la fundamentación teórica que

subyace a la creación del cuestionario y sirvió como punto de partida para identificar un banco de posibles ítems de interés (243). Este paso se realizó para asegurar que los ítems que lo forman sean representativos de las actividades habituales y que reflejen todos los aspectos claves de la AF en esta franja de edad sin omitir ninguna característica importante (245).

El procedimiento de selección y reducción de los ítems del PAQ-YC se asemejó a la propuesta de Kirshner et al. para desarrollar escalas de calidad de vida (225). En ambos casos, una vez identificadas las preguntas afines a la población, se redujo la lista a los ítems de mayor interés para presentarlos como primera propuesta. Del mismo modo, se definieron las opciones de respuesta disponibles para cada pregunta, es decir, se describió la escala asociada a cada ítem. Así mismo, el proceso de desarrollo de los ítems del PAQ-YC coincidió con las pautas de Vet et al. (201), quienes recomiendan evitar tecnicismos, palabras con múltiples significados o redacciones complejas. Tal y como indican estos autores, se excluyeron términos que llevaran a confusión a la hora de interpretar las preguntas (por ejemplo, se concretó el significado de “esfuerzo” para aclarar que se refiere a *“...todas las actividades que le hacen sudar y sentirse cansado o que hacen que se acelere su respiración como correr, saltar, hacer lanzamientos, practicar algún deporte...”*), se especificó el período de tiempo concreto al que se remonta el recuerdo (*“En los últimos 7 días...”*), se preguntó por la frecuencia y duración de las actividades utilizando un lenguaje sencillo (*“...¿qué días y durante cuánto tiempo?...”*), se revisó que los ítems estuvieran expresados de la forma más clara posible (por ejemplo, evitando formular las preguntas en negativo para no crear confusión en las respuestas) y se procuró redactar el contenido de las preguntas en un lenguaje simple que pudiera ser entendido por todas las familias, independientemente de su nivel de estudios.

Otro punto importante en la formulación inicial de los ítems del PAQ-YC fue la decisión de cuantificar la AF en términos de tipo de actividad, duración, frecuencia e intensidad y establecer los criterios para recoger la información

de estas dimensiones. La estructuración de esta parte del cuestionario presentó una importante complejidad debido a la naturaleza esporádica de la AF de los niños que hace que estas actividades sean difíciles de recordar, cuantificar y clasificar (15).

Para los ítems del 1 al 6, se utilizó un enfoque basado en categorías de actividad para cada pregunta con una tabla de respuestas que pretende recoger información detallada acerca de la duración y frecuencia. Con respecto al tipo de actividad, se descartó ofrecer listados cerrados ya que la variedad de AF practicadas por los niños dificulta disponer de un inventario de extensión moderada sin omitir ninguna acción importante para una parte de los encuestados (193). Sin embargo, se ofrecieron ejemplos de las actividades incluidas en cada ítem con el objetivo de facilitar la recuperación de recuerdos y aclarar el tipo de actividades que se piden en cada pregunta (347). Por ejemplo, en el ítem 1 se puntualizó *“...actividades dirigidas por monitor/entrenador en horario extraescolar que requieran estar activo (por ejemplo, practicando deportes, danza...”* y en el ítem 2 se especificó *“...ha jugado activamente en casa (por ejemplo, bailando, corriendo, empujando juguetes, jugando a videojuegos activos como la Nintendo Wii...) o en instalaciones de juego interiores (por ejemplo, en una ludoteca, en una piscina de bolas...)”*.

Con respecto al formato de los cuestionarios aplicables a la misma o similar franja de edad del PAQ-YC, algunos utilizan el enfoque tipo checklist, como el NPAQ (215,337), el CLASS (205), el cuestionario de AF y CS para padres de Bringolf-Isler (339) o el EY-PAQ (224), mientras que otros prefieren clasificar la AF en categorías, como el cuestionario de AF para padres y profesores de Harro (210,303), el ACTIVITY (254), el cuestionario del SAYCARE (297), el CHMS (322) o el cuestionario SHAPES (338). Por último, los cuestionarios PAQ-C (274) y PAQ-A (290) combinan ambos formatos, ya que presentan ítems que recogen categorías de actividad y un checklist compuesto por 22 actividades clasificadas como “deportes, gimnasia o danza”.

Con respecto a los cuestionarios que clasifican la AF por categorías, al igual que el PAQ-YC, es habitual que aclaren qué tipo de actividades se engloban dentro de cada ítem. Por ejemplo, el cuestionario de AF para padres y profesores de Harro proporciona ejemplos como *“...tiempo (min) que el niño estuvo involucrado en AF de baja intensidad (sentado, caminando, etc.) en interiores”* (210,303), mientras que el CHMS aclara *“...actividad física que le deja sin aliento o más caliente de lo habitual en su tiempo libre en la escuela (por ejemplo, el almuerzo)”* o *“...cuántas horas al día su hijo utiliza un ordenador (trabajando, jugando, enviando correos electrónicos, chateando, navegando por Internet, etc.)”* (322). Con respecto al PAQ-C (208,253) y al PAQ-A (341), algunas preguntas puntualizan tipos concretos de actividades, por ejemplo *“durante las clases de Educación Física, ¿cuántas veces estuviste muy activo durante las clases: jugando intensamente, corriendo, saltando, haciendo lanzamientos?”*, mientras que para otros ítems no se proporcionan ejemplos.

En relación a la duración y frecuencia, los expertos recomiendan mejorar la especificidad de los intervalos proporcionando un conjunto estrecho de períodos cuantificables en lugar de rangos amplios o adverbios generales, por ejemplo “nunca”, “rara vez”, “a veces”, “a menudo” o “siempre” (187). Por ese motivo, la recogida de información de la duración en el PAQ-YC se expresó en intervalos reducidos que van de “nada a menos de 10 minutos”, “entre 11 y 30 minutos”, “entre 31 y 59 minutos”, “entre 1 hora y 1 hora y 29 minutos”, “entre 1 hora y media y 2 horas” y “más de 2 horas”, mientras que la frecuencia se recogió para cada día de la semana. Se decidió establecer el valor de 10 minutos como la duración máxima del primer intervalo en base a las dificultades para recordar los períodos de AF de menos de 10 minutos expresadas por los encuestados en los estudios de validez del IPAQ para adultos (203).

Para los ítems restantes, las opciones de respuesta se basaron en una escala de tres opciones. El motivo que suscitó un formato diferente para las últimas

preguntas del cuestionario estaba justificado por la periodicidad de las actividades que recoge el dominio escuela. Por un lado, el tiempo dedicado al recreo, la Educación Física y la pausa de mediodía son constantes durante todo el curso académico, por lo que la ventana de oportunidad para la AF es permanente y se puede estimar. Por otro lado, la obligatoriedad de la Educación Física representa que, durante las horas semanales dedicadas a esta asignatura, se puede estimar el tiempo en que los niños están activos dentro de unos rangos de intensidad concretos. En definitiva, los patrones de AF durante las actividades de la escuela y la pausa a mediodía se desarrollan como conductas habituales y presentan menor variabilidad en relación a la AF en otros contextos, por ejemplo, el tiempo libre (68,82,83).

Los cuestionarios dirigidos a la misma o similar franja de edad que el PAQ-YC difieren en el formato y grado de detalle para recoger la duración y frecuencia de la AF. Por ejemplo, el NPAQ, en su versión original, no recoge ni la duración ni la frecuencia, sino que evalúa el grado de afinidad con cada AF del checklist en una escala Likert de 1 a 5 (337), mientras que en su versión modificada se incluye un ítem que pregunta específicamente por el promedio de tiempo (horas/día) que el niño pasa viendo la televisión (215); el cuestionario para padres y profesores de Harro (210,303), el cuestionario del SAYCARE (297) y el CLASS (205) recogen la duración y frecuencia para cada actividad; el ACTIVITY divide el día en seis franjas y registra las actividades que se realizan en cada período (254); en el SHAPES las respuestas se proporcionan indicando el número de horas (0 a 4 h) y los incrementos de 15 minutos para cada día de la semana (338), mientras que el CHMS (322) y el cuestionario de para padres de Bringolf-Isler (339) no preguntan por la frecuencia pero piden información sobre la duración de cada actividad. Además, en el cuestionario de Bringolf-Isler las opciones se presentan con intervalos de duración similares al PAQ-YC (nunca, menos de 30 minutos, de 30 minutos a 1 hora, de 1 a 2 horas, de 2 a 4 horas y más de 4 horas) (339).

Por su parte, la forma de recoger la duración de la AF en el EY-PAQ (224) es también muy parecida al formato del bloque 1 del PAQ-YC ya que, en ambos casos, las opciones de respuesta se dividen en intervalos de tiempo (excepto para duraciones superiores a 1 hora en el EY-PAQ y a 2 horas el PAQ-YC). Sin embargo, la forma de cuantificar la frecuencia es muy diferente. En el EY-PAQ se pregunta por el número de días a la semana que se desempeña cada actividad, mientras que para el PAQ-YC se recoge específicamente cada día de la semana. Con respecto al PAQ-C (208,253) y al PAQ-A (341) cabe destacar que en ningún ítem se pide información sobre la duración y únicamente se registra la frecuencia. En este contexto, el formato tipo cuadrícula del ítem 9 del PAQ-C o el ítem 8 del PAQ-A es muy similar al diseño de las opciones de respuesta del bloque 1 del PAQ-YC.

En relación a la orientación para seleccionar el valor superior o inferior del rango en los cuestionarios que ofrecen intervalos, el EY-PAQ (224) utiliza el mismo enfoque que el PAQ-YC al seleccionar el valor superior para los ítems que preguntan por CS y el valor inferior para las preguntas que se refieren a la AF. Esta decisión se tomó de acuerdo a la experiencia de los investigadores en el ámbito de la AF autoinformada, ya que se ha descrito que las personas tienden a sobreestimar las AF y a infraestimar los CS (224,295). Por el contrario, en el cuestionario de AF y CS para padres de Bringolf-Isler los intervalos se transformaron en una variable continua cogiendo el valor medio del rango para calcular el promedio (339). En este sentido, el estudio de Bringolf-Isler en el año 2010 determinó que no hay cambios significativos cuando se utiliza el valor medio en lugar del rango superior o inferior del intervalo (348).

Con respecto a la intensidad, para todos los ítems se decidió promediar el coste energético por categoría de actividad según el compendio de AF para niños y jóvenes (117). Así, se ofreció una tasación de la intensidad en términos absolutos que justifica la puntuación asignada en función de la opción de respuesta señalada por el respondedor. Se descartó recoger la

intensidad en términos relativos porque la percepción del esfuerzo depende de las características individuales (por ejemplo, la aptitud física) y condiciona que una misma actividad pueda ser descrita en diferentes grados de intensidad según el individuo (213). Así mismo, se creyó inapropiado expresar la intensidad exclusivamente en parámetros relacionados con el esfuerzo, como la sudoración, el aumento de la FC o la disnea, ya que es probable que la respuesta fisiológica a una intensidad dada sea mayor para un individuo con una mala aptitud física en comparación con sujetos con mejor una condición (120,349). Además, se ha descrito que las personas pueden definir la intensidad según diferentes señales fisiológicas, entre las que se encuentra la sudoración, el aumento de la FC, la disnea, la fatiga o el dolor muscular (187). Por tanto, y aunque se reconoce que el error de medición ocurre en todos los casos, la literatura científica muestra que se minimiza si se describe la intensidad en términos absolutos que pueden generalizarse a toda la población (100,187). Por ese motivo, la intensidad de cada ítem se estandarizó según los costes MET asociados a la categoría (117), aunque en los ítems 8, 9 y 10 se incluyó el concepto “esfuerzo físico” como complemento para puntualizar la redacción de las opciones de respuesta.

Sin embargo, cabe destacar que este enfoque no pretende aportar una equivalencia entre la puntuación del cuestionario y el GE asociado, sino que su intención es establecer una base sólida que justifique la asignación de la puntuación asociada a cada ítem. Esta decisión se tomó por varios motivos. En primer lugar, la variedad de AF incluidas en cada categoría de actividad imposibilita discernir el tiempo real dedicado a cada una de ellas. En segundo lugar, debido a la naturaleza intermitente de las AF de los niños se deduce que el tiempo señalado por los respondedores para cada ítem probablemente no esté representado íntegramente por la ejecución de dicha AF, sino que se espera que se presente combinado con períodos de descanso, especialmente para aquellas actividades de alta intensidad (15,56). En tercer lugar, el cálculo

del GE a través de una puntuación adimensional puede aumentar potencialmente el error de medición de la estimación.

Para los cuestionarios comentados anteriormente, la forma de recoger la intensidad varía en función del formato. Por ejemplo, el NPAQ no recoge información acerca de la intensidad de las actividades del checklist (215,337); en el cuestionario de AF para padres y profesores de Harro se informa por separado de las actividades de baja a moderada intensidad y de las actividades de moderada a vigorosa intensidad (210,303); el cuestionario SHAPES define la categoría AF de intensidad vigorosa como “*trotar, deportes de equipo, baile rápido, saltar a la cuerda o cualquier otra actividad física que aumente su FC y le haga respirar fuerte y sudar*” y aclara que la AF de intensidad moderada incluye “*actividades físicas de menor intensidad como caminar, ir en bicicleta a la escuela y nadar con fines recreativos*” (338); el cuestionario del SAYCARE clasifica los niveles de fatiga subjetiva en intensidad leve (“*no sudo y mi respiración es normal*”), intensidad moderada (“*estoy un poco sudoroso y con respiración rápida*”) e intensidad vigorosa (“*estoy bastante sudoroso y sin aliento*”) (297); el ACTIVITY utiliza videoclips de 3 a 5 segundos que muestran las opciones de actividad según el nivel de intensidad (sin movimiento, en movimiento y en movimiento rápido) para facilitar que el niño identifique cuál corresponde a su comportamiento (254); el CLASS (205) clasifica las actividades del checklist en moderadas o vigorosas asignando valores MET de acuerdo con los valores publicados en el compendio de AF de adultos (102), a diferencia del PAQ-YC que utiliza las referencias del compendio para jóvenes (119); mientras que en el cuestionario de AF y CS para padres de Bringolf-Isler (339) y en el CHMS (322) toda la AF se engloba dentro de la categoría moderada-vigorosa.

En relación a la estructura del cuestionario, la división en dos bloques y la categorización de los ítems por dominio y/o tipo de actividad concuerda con un estudio sobre entrevistas cognitivas en el que se resaltó la importancia de aclarar en qué contexto se desarrolla la actividad o grupo de actividades por

las que se está preguntando para evitar confusión o repetir información ya proporcionada en un ítem anterior (187). Por ese motivo, también se decidió añadir información al redactado en el enunciado de los ítems que podían presentar confusión con respecto a las actividades que se pretenden recoger. Por ejemplo, en el ítem 3 se puntualizó *“No tenga en cuenta las actividades extraescolares dirigidas que se han incluido en la pregunta 1 ni el tiempo de recreo en la escuela”*. Así mismo, se incorporó un subtítulo a cada conjunto de ítems según el dominio y/o el tipo de actividad al que hacen referencia. Los ítems 1, 2, 3 se agruparon dentro de la categoría “actividad física en el tiempo libre”; el ítem 4 se identificó como “transporte activo”; los ítems 5 y 6 se recogieron bajo el subtítulo “comportamiento sedentario en el tiempo libre”; los ítems 7, 8 y 9 se agruparon como “actividad física en la escuela” y el ítem 10 se distinguió dentro de la categoría “pausa a mediodía”. Por último, se elaboraron unas instrucciones introductorias que pusieran en contexto el objetivo del cuestionario y cautivaran el interés de los respondedores para rellenar las preguntas con la máxima atención.

Con respecto a otros cuestionarios para la franja de pediatría, en algunos casos la estructura también se ha diseñado para facilitar el recuerdo de actividades anteriores. Por ejemplo, el cuestionario para padres y profesores de Harro divide el día en dos períodos de tiempo (desde la llegada al jardín de infancia hasta el almuerzo y desde el almuerzo hasta la salida) (210,303), mientras que el ACTIVITY fracciona el día en seis etapas (de 7:00 a 8:30, de 8:32 a 9:00, de 10:00 a 10:40, de 12:00 a 12:55, de 15:15 a 16:30 y de 18:00 a 19:00) (254).

Una vez diseñada la primera versión del PAQ-YC, se inició el estudio de validez de contenido mediante un doble proceso de consulta con el objetivo de refinar el cuestionario, identificar nuevos aspectos que pudieran contribuir a la representación del constructo y mejorar su aplicabilidad (243). Si bien la validez de contenido se considera la propiedad de medición más importante

en cualquier medida indirecta, es también la más difícil de desarrollar y evaluar (294).

Los procesos de consulta a expertos y a muestras representativas de la población diana son relevantes en la creación de cualquier cuestionario, pero tienen una mayor importancia cuando el instrumento en vías de desarrollo está basado en un modelo formativo, como es el caso de los cuestionarios de AF. En los instrumentos de tipo reflexivo, los ítems son manifestaciones del constructo, lo que implica que las preguntas se correlacionan entre sí y, en algunos casos, incluso pueden omitirse o reemplazarse. Sin embargo, en los modelos formativos, cada ítem seleccionado aporta una pequeña parte del constructo que se suma a la información que proporcionan los otros ítems. Además, como las preguntas no se correlacionan necesariamente ni se realizan análisis de reducción, pasar por alto algún ítem implicaría omitir información de parte del constructo (201).

El proceso de consulta a expertos se realizó siguiendo la metodología Delphi. Este método pretende obtener una visión colectiva sobre un tema a partir de rondas repetidas de preguntas, siendo capaz de depurar las opiniones de un grupo donde el consenso se obtiene por un procedimiento de agregación de juicios individuales (326,350). Es de especial utilidad cuando hay una falta de evidencia empírica, cuando el motivo de consulta no se presta para el uso de una técnica analítica precisa o cuando se desea mantener una cierta heterogeneidad de los participantes a fin de asegurar la validez de los resultados y evitar los efectos de grupo de dominación por personalidades (326,351). En el diseño del PAQ-YC, la ausencia de un cuestionario para la franja de edad de 5 a 7 años y la necesidad de incorporar un perfil mixto de expertos para su valoración suscitó la necesidad de recurrir al método Delphi.

El procedimiento de selección de los expertos pasa necesariamente por establecer algún mecanismo de adecuación del experto con el fenómeno que se va a analizar. Esto constituye el eje central del método y clásicamente se

ha solventado de diferentes formas. Por ejemplo, Cabero Almenara propone seleccionar el panel de expertos apoyándose en la opinión del propio experto sobre su nivel de conocimiento acerca del problema analizado, así como las fuentes que le permitan argumentar la respuesta ofrecida (326,352). Por su parte, el grupo COSMIN en su estudio Delphi internacional, establecen como criterio de experticia tener al menos 5 publicaciones disponibles en PubMed sobre el objeto de consulta (294). En nuestro estudio, decimos no basarnos en un único criterio para definir el nivel de experticia, sino que preferimos establecer varios ítems que harían una media ponderada según su importancia y, en su conjunto, aportarían información sobre la idoneidad del experto (Kc). No tuvimos en cuenta ningún criterio externo de evaluación de la capacidad del experto ya que consideramos que realizar un juicio de la experticia siendo un grupo heterogéneo podría comportar incongruencias. Así mismo, seguimos la clasificación de los resultados propuesta por Cabero Almenara, seleccionando como válidos únicamente aquellos expertos que cumplieran un mínimo de puntuación de 0,8 (alta influencia) y descartando los perfiles con puntuaciones inferiores (influencia media (0,7-0,8) o baja (< 0,7)) (326,352).

En lo que se refiere al número de expertos que conforman el panel, no hay un acuerdo unánime en su determinación. Por ejemplo, Ruiz Olabuénaga e Ispuzua señalan un número entre 10 y 30 (353); García y Fernández, entre 15 y 25 (354), mientras que otros autores apuntan que, por cuestiones de operatividad, no es adecuado que supere los 50 (355). Actualmente, se recomienda que el número de miembros del panel se defina en base a los objetivos y las condiciones en las que se desarrolla cada estudio, debiendo ser lo suficientemente amplio para permitir la diversidad de opiniones pero haciendo factible el propio proceso (356). En este estudio se consideró un número entre 10 y 15 expertos siempre y cuando se pudiera asegurar la heterogeneidad necesaria para abordar el objetivo de la consulta a través de

la composición del grupo (fisioterapeutas, titulados en CAFE y maestros de Educación Infantil y Primaria).

En cuanto a la estadística durante el proceso Delphi, se ofreció en cada ronda de consulta la información sobre los resultados del grupo de cada ítem del cuestionario y se permitió que fueran objeto de debate en todas las rondas (excepto que se recomendara eliminar el ítem) con el fin de ofrecer a los miembros del panel la oportunidad de reconsiderar y, si corresponde, cambiar de opinión en las sucesivas rondas (356). Por otro lado, para definir el grado de acuerdo entre las opiniones de los participantes, se decidió emplear un doble criterio determinado por el RIR y/o el CV, tal y como apuntan en su estudio Reguant-Álvarez y Torrado-Fonseca, quienes recomiendan establecer un punto arbitrario para decidir a partir de qué valor será aceptable el consenso (327), en este caso, menor o igual al 20%. Además, se evidenció la tendencia a la convergencia de las respuestas, no observándose divergencia en ninguno de los ítems encuestados entre la primera y la segunda ronda (350).

Con respecto a la propuesta del PAQ-YC inicial, la consulta Delphi sirvió para introducir modificaciones de redactado, suprimir e incorporar ítems (ítem 7 del bloque 1 (ronda 1) e ítem 7 del bloque 2 (ronda 2), respectivamente), mejorar la claridad del contenido y, en definitiva, adaptar el cuestionario a la población diana en cuanto a su relevancia y exhaustividad. La aportación del panel de expertos permitió determinar una parte de la validez de contenido del cuestionario mediante el empleo de un doble análisis cualitativo y cuantitativo de las respuestas aportadas en las hojas de registro.

A continuación, la versión del PAQ-YC consensuada con el panel de expertos se sometió a un proceso de entrevistas cognitivas en una muestra representativa de la población diana. Tal y como apunta la evidencia científica, el estudio con las familias incluyó participantes con diferentes perfiles que garantizaran cubrir la amplitud del constructo de interés

(243,357). Concretamente, se aseguró la heterogeneidad de la muestra recogida mediante un procedimiento de muestreo estratégico no aleatorio cuyo objetivo fue maximizar la variedad de sujetos implicados en el proceso con respecto a las características diferenciadoras de la población (347). Así, deliberadamente, se contó con la participación de familias con hijos/as de 5, de 6 y de 7 años, con representación equitativa de ambos sexos. Del mismo modo, se tuvo en cuenta que participaran tanto padres como madres, de distintas edades, con diferentes niveles de estudios y pertenecientes a entornos demográficos heterogéneos.

En relación al número de entrevistados, se recomienda que el tamaño de muestra se establezca en función de la variabilidad de las características de la población y los aspectos claves del constructo (347). Algunos autores sugieren que es suficiente entre 7 y 10 entrevistas para confirmar la comprensión del tema por parte de los participantes (336,358), mientras que otros defienden que es necesario una mayor muestra (359). Debido a la falta de consenso entre los expertos, se priorizó el principio de saturación antes que el número de entrevistados, aunque se decidió reclutar al menos 7 participantes. El principio de saturación sugiere que se continúen las entrevistas hasta que la recopilación de datos cualitativos no obtenga propuestas o modificaciones relevantes (347,360). Se tuvo en cuenta este parámetro por la especial importancia de los procedimientos de entrevista para la exhaustividad y la relevancia de los ítems en los cuestionarios de tipo formativo (201). Sin embargo, se podría decir que se alcanzó la saturación desde el inicio ya que ninguno de los encuestados aportó modificaciones sustanciales. Si no se hubiera alcanzado la saturación de forma tan precoz, las entrevistas se habrían alargado hasta que las familias no aportaran nueva información, independientemente del número de sujetos entrevistados.

Con respecto a la capacitación del entrevistador, la iniciativa COSMIN recomienda que el proceso esté guiado por investigadores formados y con experiencia para estandarizar el proceso y garantizar que la información

obtenida sea relevante para el objetivo de la consulta (243). En este sentido, el entrevistador implicado en el procedimiento de validación cognitiva (MA) cumplía los requisitos de experticia recomendados por la iniciativa COSMIN respecto a la experiencia previa en la validación de métodos indirectos en pediatría (268,269) y la formación relativa al constructo y la población de interés.

En relación al método de recopilación de la información cualitativa durante las entrevistas, la evidencia científica sugiere que no hay un único enfoque válido. Los expertos recomiendan emplear un sistema que contemple el contacto directo con las familias para facilitar la interacción personal e introducir preguntas abiertas que permitan informes espontáneos. Por ejemplo, se consideran adecuados los grupos focales o las entrevistas, siempre adaptados al constructo de interés y la población de estudio (243,347). Para el estudio del PAQ-YC se seleccionó la entrevista semiestructurada como método de recopilación de datos cualitativos y se siguió el mismo guion planteado en la consulta Delphi para garantizar un enfoque coherente y ordenado, pero concediendo cierta flexibilidad para adaptar la entrevista a las inquietudes de las familias (347). La elección de la entrevista semiestructurada se justificó por la necesidad de explorar el tema de estudio de profundidad, por delante de los grupos focales, cuya principal fortaleza es su capacidad para identificar una variedad de experiencias y perspectivas interindividuales. Al tratarse de un cuestionario genérico en el que la población diana no está afectada por ningún tipo de patología, los grupos focales pierden importancia a favor de la especificidad de la entrevista. Además, se ha documentado que algunos individuos pueden mostrarse reacios a expresar su punto de vista o discutir ciertas opiniones frente a un grupo, por lo que podría perderse información por parte de participantes más introvertidos (347).

El enfoque utilizado para interrogar a los pacientes fue el sondeo verbal. En este caso, la reflexión se produce después de que el participante haya

respondido el cuestionario, por lo que evita que el proceso de interrogatorio afecte a las respuestas. Además, se ha reconocido que a través de este método el entrevistador puede obtener cierta información que de otra forma el participante no ofrecería de forma espontánea. En contraste con el sondeo verbal, la técnica de pensar en voz alta consiste en analizar los procesos de pensamiento durante la administración real del cuestionario. Dado que el rol del entrevistador se limita a leer las preguntas de la encuesta, los defensores de este procedimiento consideran que minimiza la influencia externa y puede reflejar con mayor precisión la reflexión del participante. Sin embargo, se ha descrito que algunos entrevistados experimentan dificultades para explicar sus procesos cognitivos, pueden desviarse del tema de interés o incluso pueden modificar sus respuestas por el mero hecho de verbalizar el pensamiento (347,361). Si bien ambos enfoques son válidos, algunos autores consideran que la técnica de pensar en voz alta es más apropiada para situaciones que requieren una resolución compleja de problemas, mientras que el sondeo verbal está indicado para evaluar la familiaridad del participante con el tema y la terminología planteada en cada pregunta (358). Por ello, el sondeo verbal fue el método de elección.

En relación a la recopilación de datos, aunque algunos autores recomiendan transcribir o grabar las entrevistas, en el estudio de validación cognitiva para el PAQ-YC se decidió únicamente tomar notas. Esta decisión estuvo motivada por tres razones. En primer lugar, una familia expresó incomodidad al ser grabada la conversación y se descartó este método para evitar cohibir su colaboración. En este sentido, la guía COSMIN para evaluar la validez de contenido recomienda tomar notas cuando los participantes no se encuentran cómodos con las grabaciones (243). En segundo lugar, los métodos de transcripción literal están pensados para documentar datos en entrevistas más complejas, habitualmente practicadas en contextos de grupos focales a varios participantes de forma simultánea, por lo que es fácil que se rompa el guion preestablecido y es más complejo que las notas puedan reflejar toda la

información tratada (362). Sin embargo, las entrevistas para el PAQ-YC se realizaron de forma individual, por lo que el proceso fue más sencillo y las anotaciones que se tomaron fueron suficientes para cubrir la totalidad de los aspectos comentados. En tercer lugar, la grabación puede plantear problemas de protección de datos más complejos relacionados con la identificación de los sujetos (347).

Por último, para analizar los datos recopilados durante las entrevistas, algunos autores recomiendan métodos complejos como el análisis deductivo, el análisis de marco o la codificación axial pero, de nuevo, estos métodos están pensados para procedimientos más complejos, con un mayor número de participantes y un mayor flujo de información a manejar (362). En su lugar, de acuerdo con las aportaciones de Brod et al. (347), el análisis de datos se realizó de forma iterativa, comenzando tan pronto como se obtuvieron las primeras opiniones sobre el cuestionario. De este modo, los temas emergentes identificados en el análisis de datos en curso se incorporaron en las siguientes entrevistas. Además, tal y como se ha comentado en el apartado de resultados, las aportaciones de las familias no implicaron cambios sustanciales del cuestionario y su sencillez permitió hacer un análisis y discusión cualitativas sin necesidad de recurrir a métodos complejos. Por otro lado, la participación de un número moderado de individuos y la coherencia de las opiniones con respecto al PAQ-YC facilitaron el proceso. Así mismo, se descartó la necesidad de recurrir a dos revisores que evaluaran el contenido de las entrevistas de forma independiente.

Una vez finalizado el proceso de entrevistas se dio por finalizado el estudio de validez de contenido del PAQ-YC (363). Como las aportaciones de la población diana fueron poco relevantes, se incluyeron los aspectos comentados sin necesidad de reevaluar la versión final. Esta decisión se tomó en base a las recomendaciones de la guía COSMIN, que permite introducir ajustes menores sin necesidad de volver a emitir una consulta formal (243). En relación a la falta de aportaciones mayores al cuestionario, creemos que fue debida a la

exhaustiva fase de obtención de conceptos en las primeras fases de creación del cuestionario. En este sentido, la iniciativa COSMIN reconoce que el riesgo de que se pasen por alto conceptos importantes es bajo si la fase previa a la reducción de ítems y la consulta con expertos se realiza con detenimiento (243).

La prueba de campo confirmó el proceso de obtención y reducción de ítems ya probado través de la consulta a expertos y a muestra representativas de la población diana. Según de Vet et al. las pruebas de campo forman parte de la fase de desarrollo de un instrumento y permiten detectar ciertas carencias del instrumento antes de su implementación como herramienta de medición de resultados. Por ello, lo consideran un proceso de evaluación y adaptación que debe completarse antes de que un cuestionario se aplique en investigación o práctica clínica. Sin embargo, reconocen dos particularidades de las pruebas de campo. En primer lugar, son especialmente relevantes cuando los procesos de entrevistas con expertos y con muestra representativas de la población son poco estrictos o incluso inexistentes (201). En este sentido, el desarrollo del PAQ-YC contó con participación de ambos perfiles para la adecuación de los ítems y las características generales del cuestionario, por lo que se entiende que la prueba de campo probablemente no revele la necesidad de introducir nuevos cambios. En segundo lugar, el enfoque de las pruebas de campo está muy orientado a cuestionarios contruidos bajo un modelo reflexivo, en el que la aplicación de métodos estadísticos permite discernir si todos los ítems tienen una representación en la puntuación total (201). En nuestro caso, al tratar con un cuestionario de tipo formativo no se pudieron aplicar métodos estadísticos aunque se hizo un análisis cualitativo de la contribución de los ítems a la puntuación total, observándose que todos ellos son representativos del constructo, tal y como habían apuntado los expertos y las familias que participaron en las entrevistas cognitivas. Así mismo, la ausencia de efectos techo y suelo reforzó la distribución de los ítems y demuestra la capacidad del

instrumento para detectar la diversidad de los niveles de actividad practicados por los niños de 5 a 7 años.

En relación al tamaño de la muestra, la mayor parte de las recomendaciones publicadas hacen referencia a los cálculos propios de modelos reflexivos, por lo que no son aplicables en este contexto. Cuando no es posible calcular el tamaño muestral en base a estos cálculos, la regla más utilizada utiliza la relación entre el número de sujetos y el número de ítems que forman el cuestionario. Las recomendaciones en la literatura varían de 3 a 10 observaciones para cada ítem (328,364–366), por lo que consideramos que el tamaño de muestra seleccionado es suficiente en base a las sugerencias de los expertos.

Por último, cabe mencionar que la prueba de campo fue realizada en una situación excepcional debido a la pandemia mundial del Covid-19. Por ello, entendemos que los niveles de AF reportados por este cuestionario probablemente puedan diferir de los patrones habituales y recomendamos que no se tomen estos valores como representativos de otros períodos diferentes al actual.

4.7.2. Fiabilidad y error de medición

Los procesos de comprobación de la capacidad de un cuestionario para medir el nivel de AF incluyen analizar el objetivo de evaluación y determinar cuáles son las propiedades de medición prioritarias en su validación. En el caso de los cuestionarios con fines discriminativos, como el PAQ-YC, los expertos apuntan que es imprescindible estudiar las propiedades del dominio fiabilidad (201). Por ello, una vez concluidos los estudios de desarrollo y validez de contenido se iniciaron los estudios de fiabilidad y error de medición. No se tuvo en cuenta la consistencia interna porque es una propiedad de medición irrelevante en los cuestionarios construidos bajo un modelo formativo (9,192).

El planteamiento metodológico de los estudios de fiabilidad y error de medición cumplió con los estándares sobre requerimientos generales de diseño y métodos estadísticos recogidos en el Checklist de Riesgo de Sesgo de COSMIN (250). La estabilidad de los niveles de AF entre ambas administraciones se resolvió al pedir a los participantes que respondieran el cuestionario haciendo referencia a los mismos 7 días. En relación al intervalo entre el test y el retest, se determinó un período de 1 a 3 días en base a las recomendaciones de Chinapaw et al. (9). Así mismo, al tratarse de un instrumento autoadministrado en el que no interviene la percepción del investigador, no se tuvo en cuenta el cegamiento del evaluador. Estas decisiones se tomaron en base a la evidencia demostrada por expertos en el estudio de las medidas indirectas, quienes apuntan que la calidad metodológica de los estudios realizados para probar las propiedades de medición puede interferir en gran medida en los resultados (177,202).

Con respecto a los resultados de fiabilidad obtenidos, para todas las variables analizadas se presentan índices de concordancia que superan los criterios definidos para buenas propiedades de medición, siendo excelentes tanto para los resultados de puntuación total, como para los porcentajes de AF y CS discrecional (192,247,330).

En este contexto, los cuestionarios disponibles para la misma o similar franja de edad de interés describen índices de fiabilidad muy dispares. Por un lado, el cuestionario de AF del SAYCARE muestra resultados de fiabilidad expresados a través de coeficientes de correlación de Spearman que varían de 0,28 a 0,89, siendo suficientes únicamente para las actividades de intensidad vigorosa (297). Este ejemplo es interesante porque el cuestionario comparte algunos de los atributos cualitativos con el PAQ-YC, como el constructo que mide, la administración a través de los cuidadores principales o el período de recuerdo de los últimos 7 días. Sin embargo, se diferencia en otras características, como el formato, el número de preguntas o la forma de registrar la intensidad a través de niveles de fatiga subjetiva. Así mismo, los

requisitos de diseño del estudio de fiabilidad del SAYCARE difieren del planteamiento en este estudio, lo que podría sesgar la comparación de los resultados de fiabilidad. Por ejemplo, la muestra fue reclutada en países de Sudamérica, no se proporciona evidencia de la estabilidad de las condiciones y los sujetos entre el test y el retest, los autores definen un intervalo entre administraciones de 15 días, utilizan un método estadístico diferente, etc. (297). Por otro lado, el CLASS para preescolares muestra resultados de fiabilidad muy diversos que van del 0,49 al 0,89 en función de si se analiza la duración o la frecuencia para las actividades de intensidad moderada, vigorosa o total (205), mientras que el cuestionario de AF y CS para padres de Bringolf-Isler presenta un ICC de 0,41 – 0,64 y de 0,45 – 0,67 para intervalos test-retest de 2 y 6 meses, respectivamente (339). Estos dos últimos ejemplos incluyen cuestionarios que engloban dos años de la franja de edad de la población diana del PAQ-YC pero que miden solo la participación en AFMV, con formato de preguntas tipo checklist y con períodos de recuerdo diferentes al PAQ-YC, por lo que los resultados de fiabilidad son menos comparables con los obtenidos en este estudio.

En relación a los índices de fiabilidad de los otros cuestionarios del grupo IPAQ para preescolares y escolares, el PAQ-YC muestra resultados superiores a los reportados por Bingham et al. para el EY-PAQ (224) y a los presentados en el meta-análisis de los estudios de fiabilidad del PAQ-C (capítulo 3.4.3.). Además, tanto el EY-PAQ como el PAQ-C miden únicamente la participación en AFMV, lo que pone en mayor valor los índices de fiabilidad observados en este estudio (202).

En relación al error de medición, se reconoce que todos los instrumentos presentan un cierto error en sus registros que, si no se controla, puede generar incertidumbre en las medidas recogidas (200). Sin embargo, esta propiedad de medición está poco o nada estudiada para la mayoría de los casos. De hecho, no se encontraron estudios que explorasen el error de medición en otros cuestionarios que se dirigen a la misma o similar franja de

edad que el PAQ-YC, por lo que no se pudieron comparar nuestros resultados con los reportados por otros autores. Sin embargo, tal y como se ha presentado en el capítulo 3.4.4, el SEM del PAQ-C varía de 0,15 a 0,36, mientras que el SEM del PAQ-YC fue de 0,61. Siendo el rango de puntuación de 1 a 5 para el PAQ-C y de 0 a 27 puntos para el PAQ-YC, podemos concluir que el SEM obtenido en este estudio fue menor al reportado para el PAQ-C. Aún así, cabe destacar que en ninguno de los casos pudo calificarse el resultado según los criterios para buenas propiedades de medición porque se desconoce la diferencia mínima significativa.

4.7.3. Validez de constructo

El siguiente paso en el análisis de la propiedades de medición del PAQ-YC fue el estudio de la validez de constructo por comparación del cuestionario con un dispositivo de acelerometría Actigraph wgt3x-bt (ActiGraph Corp., Pensacola, FL).

El principio básico de la validez de constructo consiste en probar hipótesis sobre las relaciones entre las puntuaciones del cuestionario en estudio y otros instrumentos que miden constructos similares u otros subgrupos conocidos (201,247). Sin embargo, la iniciativa COSMIN recoge dos aspectos adicionales dentro de la validez de constructo: la validez estructural y la validez transcultural (239). En ambos casos, se descartó analizar estas características. Por un lado, al igual que ocurría con la consistencia interna, los cuestionarios de AF no contemplan una correlación de los ítems, por lo que los análisis de la estructura interna son irrelevantes (247). Por otro lado, el estudio de validez transcultural solo tiene sentido cuando se pretende adaptar idiomática o culturalmente un cuestionario a una población diferente a la original (249).

En el caso de los cuestionarios de AF, el análisis de la validez de constructo sustituye al estudio de la validez de criterio ya que, tal y como se ha comentado en capítulos anteriores, no existe un *gold standard* real (20,111,177,190–192). Según Rennie et al. uno de los principales desafíos en los estudios de validación de los cuestionarios de AF es seleccionar un método de comparación apropiado en relación a las propiedades de medición y los resultados reportados (202). En este sentido, los acelerómetros son los sensores de movimiento más comúnmente utilizados y reconocidos como los métodos de elección para estudiar la validez de constructo de los cuestionarios que miden la AF total (13,192). Concretamente, el modelo utilizado en este estudio está considerado uno de los acelerómetros más sofisticados del mercado y cuenta con la aprobación como dispositivo médico de clase II por la Agencia de Alimentos y Medicamentos de EE.UU. (FDA) y de clase I dentro de la Unión Europea (UE) (331).

Otro de los desafíos cuando se trabaja con dispositivos de acelerometría es administrar los datos faltantes, es decir, discernir si los períodos sin registros de aceleraciones responden a un período sostenido de inactividad o son fruto de la no utilización del dispositivo. Esta situación ha llevado a los investigadores a aplicar una amplia gama de criterios para detectar el tiempo sin uso del acelerómetro, que habitualmente comprenden un número arbitrario de conteos consecutivos de ceros. Sin embargo, la aplicación de estas reglas tiene un alto riesgo de subestimar el tiempo de inactividad y puede proporcionar una evaluación inexacta del cumplimiento de los participantes (367). Por ello, autores como Winkler et al. resaltan la importancia de estandarizar los algoritmos para la detección del tiempo sin uso (368).

En este contexto, el software Actlife incluye por defecto el algoritmo sobre el tiempo de uso derivado de las investigaciones de Troiano et al. en el 2007 (369) y Choi et al. en el 2011 (332). El primero limita el tiempo de no uso del

dispositivo a 60 minutos, permitiendo recuentos con un umbral pico de 100 cpm y una tolerancia a los valores distintos de cero de 2 minutos para detectar artefactos, mientras que el segundo elimina el umbral de 100 cpm y establece el tiempo sin uso en 90 minutos, respetando la tolerancia de 2 minutos siempre y cuando los datos de los 30 minutos previos o posteriores sean recuentos de ceros consecutivos. Por un lado, el algoritmo de Troiano et al. se basa únicamente en el conjunto de datos de 2003-2004 *National Health and Nutrition Examination Survey, National Cancer Institute* para adultos (369). Por otro lado, la propuesta de Choi et al. incluye una población mixta (10-67 años) y está validada en el entorno del laboratorio pero no está demostrada en condiciones de vida libre (332).

En niños y jóvenes, las definiciones de tiempo sin uso varían de 10 a 180 min de minutos consecutivos de conteos de cero (370). En el 2015, Janssen et al. compararon los criterios de 10, 20 y 60 min de recuentos de cero consecutivos en una muestra de 957 niños de 8 a 13 años y recomendaron que se utilizara el algoritmo de 20 minutos en estudios futuros (371). Sin embargo, se ha sugerido que este umbral provoca un número poco realista de períodos sin uso, lo que llevó a Chinapaw et al. a proponer que se utilice un criterio de 60 minutos (372). En el 2018 Aadland et al. analizaron datos de acelerometría de 5203 días de 891 niños de 11 años y sugirieron aplicar un criterio de 45 o 60 minutos que no permita la tolerancia en los futuros estudios en población pediátrica. Además, sus hallazgos suscitan que los períodos de tiempo sin uso se deben incrementar a medida que aumenta la edad, ya que es más frecuente que pasen períodos de tiempo cada vez más largos en actividades sedentarias (373). Sin embargo, la base de datos internacional de acelerometría infantil (ICAD), aplica un criterio de 60 minutos con tolerancia de 2 minutos de recuentos distintos de cero (374). Por ello, el algoritmo de tiempo sin uso seleccionado en este estudio se basó en la propuesta de Choi et al. (332) pero adaptada al registro de 60 minutos consecutivos de ceros para cumplir con las recomendaciones en el franja de pediatría. Además, se consideró permitir

2 minutos de tolerancia pero con el requisito de que los 30 minutos previos y posteriores fuera también recuentos de ceros para evitar infraestimar el tiempo dedicado a los CS. Sin embargo, cabe mencionar que la ausencia de un algoritmo estandarizado para todos los estudios pueda influir en gran medida en las estimaciones reportadas. De hecho, se ha documentado que el enfoque de análisis para determinar el tiempo sin uso puede condicionar diferencias significativas en los resultados (375).

En relación al cumplimiento de uso del dispositivo para considerar un día válido, las recomendaciones varían entre 8 y 10 horas, aunque los expertos apuntan que los mejores resultados se obtienen a partir de registros de 10 o más horas (370,372,376). Con respecto a la elección del intervalo para agrupar los datos brutos, Chinapaw et al. sugieren utilizar períodos de época pequeños que compendien los datos en ciclos reducidos, preferiblemente iguales o inferiores a 15 segundos (9). Estas recomendaciones fueron consideradas en el diseño de este estudio con la intención de que los resultados fueran analizados bajo un enfoque lo más adaptado posible a las particularidades de la población infantil.

En relación a la categorización de los datos brutos de salida se han descrito ecuaciones de regresión que estiman el tiempo dedicado a los diferentes niveles de intensidad. Sin embargo, tampoco existe consenso sobre los puntos de corte apropiados y los criterios utilizados incluyen una nueva fuente de error (174,184,185). Además, dado que los dispositivos basados en acelerometría utilizan un procedimiento integrador para resumir los recuentos de movimiento, el valor de los cpm puede ocultar períodos cortos de actividad vigorosa si se alterna con períodos de descanso (56).

En este contexto, el software Actilife ofrece un total de 19 posibles algoritmos, de los cuales 6 fueron diseñados para ser utilizados en muestras de participantes entre 6 y 18 años (316,323,324,377-379); 5 para niños entre 3 y 5 años (380,381); y 1 para niños de 1 a 2 años (382). Para este estudio

seleccionamos el algoritmo de Evenson et al. (323) para niños de 5 a 8 años porque la muestra de participantes reclutada por estos autores era muy similar en edad a la población del PAQ-YC. Sin embargo, cabe destacar que este algoritmo utiliza únicamente los recuentos del eje vertical para estimar los porcentajes en cada nivel de intensidad. Por tanto, se suma una nueva fuente de error que podría explicar las diferencias entre los resultados del cuestionario y el acelerómetro.

Así mismo, el software ofrece la posibilidad de categorizar los datos en unidades de GE. Sin embargo, se descartó porque los algoritmos destinados a este cálculo están diseñados específicamente para adultos (319,383,384) y porque la intención del PAQ-YC no es ofrecer estimaciones del GE. Con respecto al primer punto, se recomienda utilizar esta opción únicamente cuando los participantes son mayores de 19 años, ya que los cálculos no tienen en cuenta la TMR (385).

En relación a los resultados obtenidos para la validez de constructo, en todos los casos se presentan coeficientes de correlación suficientes en base a los criterios definidos por Terwee et al (192). Estos autores defienden que la validez de constructo de un cuestionario que mide la AF total debe obtener un coeficiente de correlación de al menos 0,5 por comparación con los cpm totales de un acelerómetro. Así mismo, los resultados cumplen las recomendaciones de Pols et al. quienes sugieren que la correlación entre los cuestionarios de AF y los instrumentos comparadores debe oscilar entre 0,5 y 0,7 (177). Sin embargo, no superan el mínimo establecido por Hidding et al. para considerar que este estudio ofrece un nivel de evidencia 1, aunque el valor obtenido se aproxima al criterio de 0,6 definido por estos autores (13).

Con respecto a la validez de constructo para los otros cuestionarios de AF del grupo IPAQ para la franja de pediatría, en ambos casos, los resultados obtenidos en este estudio superan los índices de correlación reportados. Por un lado, Bingham et al. informan un coeficiente de correlación entre la AFMV

medida por el cuestionario y los recuentos de acelerometría de 0,30 para el EY-PAQ (224), mientras que el meta-análisis de los estudios de validez de constructo que pueden ofrecer un nivel de evidencia 1 para el PAQ-C mostró una correlación agrupada de 0,34 (capítulo 3.4.5).

En los cuestionarios de AF disponibles para la misma o similar franja de edad que el PAQ-YC se observan resultados dispares. Por un lado, el cuestionario de Harro original fue comparado con los resultados de un acelerómetro uniaxial y un monitor de FC, mostrando coeficientes de correlación de 0,53, y de 0,40, respectivamente (210). Por otro lado, su versión modificada presentó una correlación de 0,10 por comparación con los minutos de AFMV recogidos a través de un acelerómetro (303). Así mismo, la puntuación total del ACTIVITY mostró un coeficiente de correlación de 0,40 contra los cpm de actividad de un acelerómetro Caltrac y de 0,17 por comparación con los registros de un monitor de FC (254). Por último, la participación en AFMV estimada a través del SAYCARE presenta correlaciones de 0,44 con los recuentos de AFMV de un acelerómetro (297).

Con respecto a los cuestionarios que incluyen al menos dos años de la franja de edad de la población diana del PAQ-YC, los resultados son también desiguales. En primer lugar, el CLASS para preescolares fue comparado contra los recuentos de AFMV de un acelerómetro mostrando coeficientes de correlación de -0,06 y -0,04, respectivamente (205). En segundo lugar, el cuestionario SHAPES mostró una correlación de 0,44 contra los registros de AFMV de un acelerómetro (338). En tercer lugar, el cuestionario de AF y CS para padres de Bringolf-Isler fue comparado contra los minutos diarios de AFMV de un acelerómetro ajustado por edad, sexo, curso escolar y educación de los padres y mostró coeficientes de correlación que varían entre 0,07 y 0,21 (339). En cuarto lugar, la participación en AFMV estimada a través del CHMS y los recuentos de AFMV de acelerómetros presentan correlaciones de 0,29 (322). Por último, la Encuesta de Salud para el Cuestionario de AF de

Inglaterra mostró una correlación con los registros de un acelerómetro de 0,16 (340).

Cabe destacar que estos enfoques difieren del planteamiento metodológico del estudio de validez de constructo del PAQ-YC, tanto por el cumplimiento de los estándares de calidad metodológica como por la elección del instrumento y/o constructo de comparación. Además, cabe destacar la mayor dificultad para lograr una correlación suficiente para el PAQ-YC en comparación con otros cuestionarios que registran únicamente la AFMV, ya que es más complejo de recordar para el respondedor (202).

4.7.4. Puntos fuertes y limitaciones

Creemos que una de las mayores fortalezas del PAQ-YC es el exhaustivo análisis de la literatura científica para establecer el marco conceptual que fundamentó el procedimiento de obtención y reducción de ítems. Por otro lado, para la mayoría de los cuestionarios existe muy poca información acerca de los procesos de consulta a expertos y a muestras representativas de la población. Sin embargo, para el PAQ-YC estas explicaciones son extensas y permiten evaluar de forma concisa su validez de contenido.

Otro punto que destacar es la justificación del sistema de puntuación en base a los valores de referencia de consumo energético, las recomendaciones de AF y el tiempo dedicado a cada categoría de actividad. Habitualmente, los manuales de corrección de los cuestionarios incluyen el algoritmo de puntuación sin explicar el porqué de las puntuaciones asignadas.

Otro de los puntos fuertes de los estudios presentados sobre las propiedades de medición del PAQ-YC fue que el planteamiento metodológico cumplió con los estándares sobre requerimientos generales de diseño y métodos estadísticos recogidos en el Checklist de Riesgo de Sesgo de COSMIN (250). Si bien esta herramienta se ha diseñado específicamente para evaluar la calidad

de los estudios sobre una propiedad de medición, los estándares que recoge dan una aproximación de los criterios recomendados en este tipo de estudios.

Por un lado, para el estudio de fiabilidad y error de medición se tuvo en cuenta realizar dos administraciones independientes con la seguridad de que el intervalo entre ambas fuera apropiado en relación al período de recuerdo, a la estandarización de las condiciones de la prueba y al período de blanqueamiento entre el test y el retest (9,192,201).

Por otro lado, para el estudio de validez de constructo, los criterios de recopilación de los datos de acelerometría se establecieron en base a las recomendaciones de los expertos para ajustarse a las particularidades de la edad de la muestra del estudio. En primer lugar, la elección de un instrumento que permitió registrar la AF practicada durante el período de recuerdo al que se refiere el cuestionario aseguró la monitorización del participante durante la semana que posteriormente tuvo que informar (219). En segundo lugar, se tuvo en cuenta que el uso de la acelerometría debía presentar un alto cumplimiento por parte de los participantes y debía asegurar la representación de los días entre semana y el fin de semana, ya que se ha descrito que los niveles de actividad pueden variar considerablemente (172). En tercer lugar, se insistió en la educación de los adultos responsables sobre cómo y cuándo usar el monitor de actividad con materiales gráficos y demostraciones que estandarizaban el protocolo de medición (95).

Las principales limitaciones hacen referencia a los estudios de error de medición y validez de constructo.

Con respecto al primero, el desconocimiento de la diferencia mínima significativa y falta de estudios que aporten evidencia sobre el error de medición en la mayoría de los cuestionarios publicados dificultaron establecer conclusiones en relación a esta propiedad de medición.

Para el estudio de validez de constructo, se ponen de manifiesto limitaciones derivadas de la recopilación y del procesamiento de los datos de los acelerómetros que, en cierta medida, podrían justificar las diferencias obtenidas entre los registros del PAQ-YC y los recuentos de actividad. En primer lugar, los dispositivos de acelerometría presentan algunas limitaciones intrínsecas, como la imposibilidad de registrar ciertas actividades que no implican variaciones en la aceleración o actividades que requieren un mayor esfuerzo, como el transporte de una carga de peso (56,95). En segundo lugar, uno de los puntos débiles más importantes fue la limitación del uso de los acelerómetros durante las actividades acuáticas. Esta es una importante consideración en los estudios en pediatría, especialmente a la edad de 5 a 7 años, en la que la mayoría de los niños acuden a cursos de natación. Por ello, se deduce que los resultados de la acelerometría reportados en este estudio subestiman el volumen real de AF (95,100). En tercer lugar, la falta de estandarización de los criterios para definir el tiempo sin uso y los puntos de corte por niveles de actividad pudieron añadir una nueva fuente de error a los resultados (174,184,185). Por último, las estimaciones de los patrones de actividad probablemente sean inexactas y subestimen el nivel de AF real porque se calculan a partir de datos que, en algunos casos, no reflejan la totalidad de las horas de vigilia (386).

4.7.5. Consideraciones generales

Cada uno de los atributos cualitativos y las propiedades de medición descritas ponen de manifiesto que el PAQ-YC es un cuestionario válido y fiable para medir la AF en niños con edades comprendidas entre los 5 y los 7 años.

Por tanto, disponemos de un nuevo instrumento de bajo coste, que permite recopilar información en un gran número de sujetos, es capaz de identificar las dimensiones y los dominios en los que se desarrolla la AF y no tiene riesgo

de afectar a los patrones de AF porque el período de referencia es anterior a su administración.

Además, el equipo investigador coincide en que el PAQ-YC sea de uso libre y no esté sometido a derechos de copyright, por lo que puede ser utilizado sin restricciones y de forma gratuita.

5. Conclusiones

En base a los resultados presentados en la revisión sistemática sobre propiedades de medición de los cuestionarios del grupo International Physical Activity Questionnaire (IPAQ) para niños en edad preescolar y escolar, se extraen las siguientes conclusiones:

1. Los cuestionarios del grupo IPAQ dirigidos a población preescolar y escolar son el Early Years - Physical Activity Questionnaire (EY-PAQ) para menores de 5 años y el Physical Activity Questionnaire – Older Children (PAQ-C) para niños de 8 a 14 años.
2. No se pudo evaluar la validez de contenido, el error de medición y la capacidad de respuesta de los cuestionarios Early Years - Physical Activity Questionnaire (EY-PAQ) y Physical Activity Questionnaire – Older Children (PAQ-C).
3. La fiabilidad del Early Years - Physical Activity Questionnaire (EY-PAQ) fue insuficiente, mientras que para el Physical Activity Questionnaire – Older Children (PAQ-C) se calificó como suficiente, ambas con calidad de evidencia muy baja.
4. La validez de constructo tanto del Early Years - Physical Activity Questionnaire (EY-PAQ) como del Physical Activity Questionnaire – Older Children (PAQ-C) fue insuficiente, con calidad de la evidencia muy baja y moderada, respectivamente.
5. La revisión sistemática sobre propiedades de medición de cuestionarios del grupo International Physical Activity Questionnaire (IPAQ) dirigidos a población preescolar y escolar reveló la ausencia de un cuestionario que cubra el rango de edad de 5 a 7 años.

En base a los resultados presentados en los estudios de desarrollo y análisis de propiedades de medición del Physical Activity Questionnaire – Young Children (PAQ-YC), se extraen las siguientes conclusiones:

1. El Physical Activity Questionnaire – Young Children (PAQ-YC) muestra una validez de contenido adecuada en términos de validez aparente, relevancia, exhaustividad y comprensibilidad de los ítems.
2. No existen efectos techo ni suelo en la puntuación del Physical Activity Questionnaire – Young Children (PAQ-YC).
3. La fiabilidad test-retest del Physical Activity Questionnaire – Young Children (PAQ-YC) muestra índices de concordancia intraobservador excelentes para los resultados de puntuación.
4. La validez de constructo del Physical Activity Questionnaire – Young Children (PAQ-YC) mediante comparación de métodos con acelerometría Actigraph wGT3X-BT (ActiGraph Corp., Pensacola, FL) presenta resultados de correlación moderados para todas las variables analizadas.

6. Líneas de futuro

Durante la realización de los estudios que forman parte de esta tesis doctoral han surgido posibles líneas de investigación futuras que podrían complementar los hallazgos obtenidos.

Por un lado, creemos que es interesante implementar nuevos estudios que exploren la interpretabilidad de los resultados del PAQ-YC. En primer lugar, proponemos definir los puntos de corte en la puntuación para discriminar entre individuos físicamente activos e individuos inactivos según si cumplen o no las recomendaciones internacionales con respecto al volumen acumulado de AF (3). Este planteamiento se extrae de estudios previos que han descrito puntos de corte para clasificar a los niños de 8 a 14 años según su nivel de AF autoinformada con el PAQ-C (387–390). En segundo lugar, proponemos describir la diferencia mínima significativa del PAQ-YC, tanto por su interés en la interpretabilidad clínica de los resultados, como por su valor para calificar el error de medición según los criterios para buenas propiedades de medición (247).

Por otro lado, creemos que otra línea de investigación prometedora sería el estudio de la capacidad de respuesta del PAQ-YC para explorar su utilidad como índice evaluativo para monitorear los patrones de AF a lo largo del tiempo o para evaluar el efecto de una intervención terapéutica.

Así mismo, tenemos la intención de adaptar el PAQ-YC a poblaciones específicas. En concreto, hemos previsto estudiar las propiedades de medición del cuestionario en pacientes con cardiopatías congénitas y enfermedades cardiovasculares hereditarias. En estas poblaciones consideramos de especial interés el estudio de la capacidad de respuesta como una herramienta de valoración de los programas de rehabilitación cardíaca, sobre todo para aquellos en lista de trasplante cardíaco .

7. Referencias bibliográficas

1. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Scientific Report, 2008. Washington, DC; 2008.
2. Caspersen CJ, Powell KE CG, Siscovick DS, Laporte RE, Newman J, Health ; Iverson DC, Fielding JE, et al. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Heal Rep.* 1985;100(2):126–31.
3. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Scientific Report, 2018. Washington, DC; 2018.
4. Lalonde M. A new Perspective on the Health of Canadians. A Working Document. Ottawa: Government of Canada; 1974.
5. U.S. Department of Health, Education, and Welfare. Healthy people: The Surgeon General's report on health promotion and disease prevention. Washington, DC: United States Public Health Service; 1979.
6. Herting MM, Chu X. Exercise, cognition, and the adolescent brain. *Birth Defects Res.* 2017;109(20):1672–9.
7. Kohl HW, Fulton JE, Caspersen CJ. Assessment of physical activity among children and adolescents: A review and synthesis. *Prev Med (Baltim).* 2000;31(2 II):S54–76.
8. Tessier S, Vuillemin A, Briançon S. Review of physical activity questionnaires validated for children and adolescents. *Sci Sports.* 2008;23:118–25.
9. Chinapaw MJMM, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. *Sports Med.* 2010;40(7):539–63.

10. Foley L, Maddison R, Olds T, Ridley K. Self-report use-of-time tools for the assessment of physical activity and sedentary behaviour in young people: Systematic review. *Obes Rev.* 2012;13(8):711–22.
11. Cancela JM, Ayán C, Castro A. An evaluation of questionnaires assessing physical activity levels in youth populations. *J Child Heal Care.* 2013;17(3):274–93.
12. Martínez-Lemos I, Ayán-Pérez C, Sánchez Lastra A, Cancela Carral J, Valcarce Sánchez R. Physical activity questionnaires for Spanish children and adolescents: a systematic review. *An Sist Sanit Navar.* 2016;39(3):417–28.
13. Hidding LM, Chinapaw MJM, van Poppel MNM, Mokkink LB, Altenburg TM. An Updated Systematic Review of Childhood Physical Activity Questionnaires. *Sport Med.* 2018;48(12):2797–842.
14. Pate RR. Physical Activity Assessment in Children and Adolescents. *Crit Rev Food Sci Nutr.* 1993;33(4–5):321–6.
15. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sport Med.* 2001;31(6):439–54.
16. National Institutes of Health. Physical Activity and Cardiovascular Health. NIH Consens Statement Online. 1995.
17. U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: U.S. Government Printing Office; 1996.
18. WHO. Global action plan on physical activity 2018–2030: more active people for a healthier world. Geneva (Switzerland): World Health Organization; 2018.
19. Piggin J. What Is Physical Activity? A Holistic Definition for Teachers,

- Researchers and Policy Makers. *Front Sport Act Living*. 2020;2:72.
20. Pettee Gabriel KK, Morrow JR, Woolsey A-LT. Framework for physical activity as a complex and multidimensional behavior. *J Phys Act Health*. 2012;9(Suppl 1):S11–8.
 21. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: The population health science of sedentary behavior. *Exerc Sport Sci Rev*. 2010;38(3):105–13.
 22. Tremblay M. Letter to the editor: Standardized use of the terms “sedentary” and “sedentary behaviours.” *Appl Physiol Nutr Metab*. 2012;37(3):540–2.
 23. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. *Lancet*. 2016;388(10051):1302–10.
 24. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: Conceptual review and framework for intervention research. *Int J Behav Nutr Phys Act*. 2017;14(1):156.
 25. Roberts KC, Butler G, Branchard B, Rao DP, Otterman V, Thompson W, et al. The physical activity, sedentary behaviour and sleep (PASS) indicator framework. *Heal Promot Chronic Dis Prev Canada*. 2017;37(8):252–6.
 26. Klepac Pogrmilovic B, O’Sullivan G, Milton K, Biddle SJH, Bauman A, Bellew W, et al. The development of the Comprehensive Analysis of Policy on Physical Activity (CAPPA) framework. *Int J Behav Nutr Phys Act*. 2019;16(1):60.

27. Dobbels F, de Jong C, Drost E, Elberse J, Feridou C, Jacobs L, et al. The PROactive innovative conceptual framework on physical activity. *Eur Respir J*. 2014;44(5):1223–33.
28. Thelen E, Smith LB. A dynamic systems approach to the development of cognition and action. Cambridge. M. Editor; 1994.
29. Rowland TW. The biological basis of physical activity. *Med Sci Sports Exerc*. 1998;30(3):392–9.
30. Eisenmann JC, Wickel EE. The biological basis of physical activity in children: Revisited. *Pediatr Exerc Sci*. 2009;21(3):257–72.
31. Franks PW, Ravussin E, Hanson RL, Harper IT, Allison DB, Knowler WC, et al. Habitual physical activity in children: The role of genes and the environment. *Am J Clin Nutr*. 2005;82(4):901–8.
32. Spence J, Lee R. Toward a comprehensive model of physical activity. *Psych Sport Ex*. 2003;4(1):7–24.
33. Giles-Corti B, Timperio A, Bull F, Pikora T. Understanding physical activity environmental correlates: Increased specificity for ecological models. *Exerc Sport Sci Rev*. 2005;33(4):175–81.
34. Carson RL, Castelli DM, Beighle A, Erwin H. School-based physical activity promotion: a conceptual framework for research and practice. *Child Obes*. 2014;10(2):100–6.
35. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol*. 2000;55(1):68–78.
36. Weaver RG, Beets MW, Webster C, Beighle A, Huberty J. A Conceptual Model for Training After-School Program Staffers to Promote Physical Activity and Nutrition. *J Sch Health*. 2012;82(4):186–95.

37. Lubans DR, Lonsdale C, Cohen K, Eather N, Beauchamp MR, Morgan PJ, et al. Framework for the design and delivery of organized physical activity sessions for children and adolescents: rationale and description of the 'SAAFE' teaching principles. *Int J Behav Nutr Phys Act.* 2017;14:24.
38. Teixeira PJ, Carraça E V., Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: A systematic review. *Int J Behav Nutr Phys Act.* 2012;9(1):78.
39. Sanders GJ, Juvancic-Heltzel J, Williamson ML, Roemmich JN, Feda DM, Barkley JE. The effect of increasing autonomy through choice on young children's physical activity behavior. *J Phys Act Heal.* 2016;13(4):428–32.
40. Elhakeem A, Cooper R, Bann D, Hardy R. Childhood socioeconomic position and adult leisure-time physical activity: a systematic review. *Int J Behav Nutr Phys Act.* 2015;12(1):92.
41. Auhuber L, Vogel M, Grafe N, Kiess W, Poulain T. Leisure Activities of Healthy Children and Adolescents. *Int J Environ Res Public Health.* 2019;16(12):2078.
42. Dudley DA, Okely AD, Pearson P, Cotton WG, Caputi P. Changes in physical activity levels, lesson context, and teacher interaction during physical education in culturally and linguistically diverse Australian schools. *Int J Behav Nutr Phys Act.* 2012;9(1):114.
43. Dudley D, Cotton W, Peralta L, Winslade M. A stepped-wedge implementation and evaluation of the healthy active peaceful playgrounds for youth (HAPPY) intervention. *BMC Public Health.* 2018;18(1):532.
44. Hesketh KR, Lakshman R, van Sluijs EMF. Barriers and facilitators to

- young children's physical activity and sedentary behaviour: a systematic review and synthesis of qualitative literature. *Obes Rev.* 2017;18(9):987–1017.
45. Encuesta Nacional de Salud ENSE, España 2017. SG Información Sanitaria. Ministerio de Sanidad, Consumo y Bienestar Social; 2017.
 46. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity and cardiovascular disease. *Circulation.* 2012;125(5):729–37.
 47. Oliveira AF, Moreira C, Abreu S, Mota J, Santos R. Environmental determinants of physical activity in children: A systematic review. *Arch Exerc Heal Dis.* 2014;4(2):254–61.
 48. Smith M, Hosking J, Woodward A, Witten K, MacMillan A, Field A, et al. Systematic literature review of built environment effects on physical activity and active transport - an update and new findings on health equity. *Int J Behav Nutr Phys Act.* 2017;14(1):158.
 49. Estudio ALADINO 2019: Estudio sobre Alimentación, Actividad Física, Desarrollo Infantil y Obesidad en España 2019. Agencia Española de Seguridad Alimentaria y Nutrición. Ministerio de Consumo; 2020.
 50. Atkin AJ, Sharp SJ, Harrison F, Brage S, Van Sluijs EMF. Seasonal Variation in Children's Physical Activity and Sedentary Time. *Med Sci Sports Exerc.* 2016;48(3):449–56.
 51. Vanderloo LM, Martyniuk OJM, Tucker P. Physical and sedentary activity levels among preschoolers in home-based childcare: A systematic review. *J Phys Act Heal.* 2015;12(6):879–89.
 52. Maitland C, Stratton G, Foster S, Braham R, Rosenberg M. A place for play? The influence of the home physical environment on children's

- physical activity and sedentary behaviour. *Int J Behav Nutr Phys Act.* 2013;10:99.
53. Gray C, Gibbons R, Larouche R, Sandseter EBH, Bienenstock A, Brussoni M, et al. What is the relationship between outdoor time and physical activity, sedentary behaviour, and physical fitness in children? A systematic review. *Int J Environ Res Public Health.* 2015;12(6):6455–74.
 54. Arundell L, Fletcher E, Salmon J, Veitch J, Hinkley T. A systematic review of the prevalence of sedentary behavior during the after-school period among children aged 5-18 years. *Int J Behav Nutr Phys Act.* 2016;13(1):93.
 55. Kaushal N, Rhodes RE. The home physical environment and its relationship with physical activity and sedentary behavior: A systematic review. *Prev Med (Baltim).* 2014;67:221–37.
 56. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. *Res Q Exerc Sport.* 2000;71(Suppl 2):59.
 57. Tasas de escolarización por edad en niveles no obligatorios [Internet]. Instituto Nacional de Estadística (INE). Available from: <https://www.ine.es/jaxiT3/Datos.htm?t=12542#!tabs-tabla>
 58. Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, et al. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation.* 2013;128(20):2259–79.
 59. Howley E. Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Med Sci Sports Exerc.* 2001;33(Suppl 6):S364-420.

60. Pate RR, Pfeiffer KA, Trost SG, Ziegler P, Dowda M. Physical activity among children attending preschools. *Pediatrics*. 2004;114(5):1258–63.
61. Ross SET, Dowda M, Dishman RK, Pate RR. Classes of physical activity and sedentary behavior in 5th grade children. *Am J Health Behav*. 2016;40(3):352–61.
62. Tremblay MS, LeBlanc AG, Carson V, Choquette L, Gorber SC, Dillman C, et al. Canadian sedentary behaviour guidelines for the early years (aged 0-4 years). *Appl Physiol Nutr Metab*. 2012;37(2):370–80.
63. Tremblay MS, LeBlanc AG, Janssen I, Kho ME, Hicks A, Murumets K, et al. Canadian sedentary behaviour guidelines for children and youth. *Appl Physiol Nutr Metab*. 2011;36(1):59–64.
64. Hume C, Salmon J, Ball K. Children’s perceptions of their home and neighborhood environments, and their association with objectively measured physical activity: A qualitative and quantitative study. *Health Educ Res*. 2005;20(1):1–13.
65. Branscum P, Sharma M. After-school based obesity prevention interventions: A comprehensive review of the literature. *Int J Environ Res Public Health*. 2012;9(4):1438–57.
66. Mahoney JL, Lord H, Carryl E. Afterschool program participation and the development of child obesity and peer acceptance. *Appl Dev Sci*. 2005;9(4):202–15.
67. Halpern R. The promise of after-school programs for low-income children. *Early Child Res Q*. 2000;15(2):185–214.
68. Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria. *Boletín Oficial del Estado*

- (52) [Internet]. Ministerio de Educación, Cultura y Deporte; p. 19349-19420. 1 de marzo de 2014. Available from: <https://www.boe.es/buscar/doc.php?id=BOE-A-2014-2222>
69. Pate RR, Davis MG, Robinson TN, Stone EJ, McKenzie TL, Young JC. Promoting physical activity in children and youth: A leadership role for schools - A scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism (Physical Activity Committee). *Circulation*. 2006;114(11):1214-24.
 70. Hills AP, Dengel DR, Lubans DR. Supporting Public Health Priorities: Recommendations for Physical Education and Physical Activity Promotion in Schools. *Prog Cardiovasc Dis*. 2015;57(4):368-74.
 71. Hatfield DP hili., Chomitz VR al. Increasing Children's Physical Activity During the School Day. *Curr Obes Rep*. 2015;4(2):147-56.
 72. Kibbe DL, Hackett J, Hurley M, McFarland A, Schubert KG, Schultz A, et al. Ten Years of TAKE 10!®: Integrating physical activity with academic concepts in elementary school classrooms. *Prev Med (Baltim)*. 2011;52(Suppl 1):S43-50.
 73. Ridgers ND, Stratton G, Fairclough SJ. Physical activity levels of children during school playtime. *Sport Med*. 2006;36(4):359-71.
 74. Parrish AM, Chong KH, Moriarty AL, Batterham M, Ridgers ND. Interventions to Change School Recess Activity Levels in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sport Med*. 2020;50(12):2145-73.
 75. Escalante Y, García-Hermoso A, Backx K, Saavedra JM. Playground Designs to Increase Physical Activity Levels During School Recess: A Systematic Review. *Heal Educ Behav*. 2014;41(2):138-44.

76. Cardon G, Van Cauwenberghe E, Labarque V, Haerens L, De Bourdeaudhuij I. The contribution of preschool playground factors in explaining children's physical activity during recess. *Int J Behav Nutr Phys Act.* 2008;5:11.
77. Black IE, Menzel NN, Bungum TJ. The Relationship Among Playground Areas and Physical Activity Levels in Children. *J Pediatr Heal Care.* 2015;29(2):156-68.
78. Ridgers ND, Salmon J, Parrish AM, Stanley RM, Okely AD. Physical activity during school recess: A systematic review. *Am J Prev Med.* 2012;43(3):320-8.
79. Lonsdale C, Rosenkranz RR, Peralta LR, Bennie A, Fahey P, Lubans DR. A systematic review and meta-analysis of interventions designed to increase moderate-to-vigorous physical activity in school physical education lessons. *Prev Med (Baltim).* 2013;56(2):152-61.
80. Lorås H. The Effects of Physical Education on Motor Competence in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports.* 2020;8(6):88.
81. McKenzie TL. Physical education's role in public health. *Res Q Exerc Sport.* 1991;62(2):124-37.
82. Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado (340)* [Internet]. Jefatura del Estado; p. 122868-122953. 30 de diciembre de 2020. Available from: <https://www.boe.es/boe/dias/2020/12/30/pdfs/BOE-A-2020-17264.pdf>
83. Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado (106)* [Internet]. Jefatura del Estado; p. 17158-17207. 4 de mayo

de 2006. Available from:
<https://www.boe.es/buscar/doc.php?id=BOE-A-2006-7899>

84. Larouche R, Saunders TJ, Faulkner GEJ, Colley R, Tremblay M. Associations between active school transport and physical activity, body composition, and cardiovascular fitness: a systematic review of 68 studies. *J Phys Act Health*. 2014;11(1):206–27.
85. Faulkner GEJ, Buliung RN, Flora PK, Fusco C. Active school transport, physical activity levels and body weight of children and youth: A systematic review. *Prev Med (Baltim)*. 2009;48(1):3–8.
86. Rojas-Rueda D, De Nazelle A, Andersen ZJ, Braun-Fahrländer C, Bruha J, Bruhova-Foltynova H, et al. Health impacts of active transportation in Europe. *PLoS One*. 2016;11(3):e0149990.
87. Hills AP, Byrne NM. An overview of physical growth and maturation. *Med Sport Sci*. 2010;55:1–13.
88. Ingram DK. Age-related decline in physical activity: Generalization to nonhumans. *Med Sci Sports Exerc*. 2000;32(9):1623–9.
89. Sisk CL, Zehr JL. Pubertal hormones organize the adolescent brain and behavior. *Front Neuroendocrinol*. 2005;26(3–4):163–74.
90. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. *Med Sci Sports Exerc*. 2000;32(5):963–75.
91. Hills AP, King NA, Armstrong TP. The Contribution of Physical Activity and Sedentary Behaviours to the Growth and Development of Children and Adolescents. Implications for Overweight and Obesity. *Sport Med*. 2007;37(6):533–45.
92. Pellegrini AD, Smith PK. Physical activity play: The nature and function

- of a neglected aspect of play. *Child Dev.* 1998;69(3):577–98.
93. Dwyer GM, Baur LA, Hardy LL. The challenge of understanding and assessing physical activity in preschool-age children: Thinking beyond the framework of intensity, duration and frequency of activity. *J Sci Med Sport.* 2009;12(5):534–6.
 94. Sturgess J. A model describing play as a child-chosen activity — is this still valid in contemporary Australia? *Aust Occup Ther J.* 2003;50(2):104–8.
 95. Trost SG. Objective measurement of physical activity in youth: Current issues, future directions. *Exerc Sport Sci Rev.* 2001;29(1):32–6.
 96. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children’s physical activities: An observational study. *Med Sci Sports Exerc.* 1995;27(7):1033–41.
 97. Ashiabi GS. Play in the preschool classroom: Its socioemotional significance and the teacher’s role in play. *Early Child Educ J.* 2007;35(2):199–207.
 98. Ridgers ND, Timperio A, Crawford D, Salmon J. What Factors Are Associated with Adolescents’ School Break Time Physical Activity and Sedentary Time? *PLoS One.* 2013;8(2):e56838.
 99. Ravussin E, Bogardus C. A brief overview of human energy metabolism and its relationship to essential obesity. *Am J Clin Nutr.* 1992;55(Suppl 1):S242–5.
 100. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *Eur J Epidemiol.* 2007 Jun;22(6):353–62.
 101. Compher C, Frankenfield D, Keim N, Roth-Yousey L. Best Practice

- Methods to Apply to Measurement of Resting Metabolic Rate in Adults: A Systematic Review. *J Am Diet Assoc.* 2006;106(6):881–903.
102. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc.* 1993;25(1):71–4.
 103. Herrmann SD, McMurray RG, Kim Y, Willis EA, Kang M, McCurdy T. The influence of physical characteristics on the resting energy expenditure of youth: A meta-analysis. *Am J Hum Biol.* 2017;29(3).
 104. Harrell JS, McMurray RG, Baggett CD, Pennell ML, Pearce PF, Bangdiwala SI. Energy costs of physical activities in children and adolescents. *Med Sci Sports Exerc.* 2005;37(2):329–36.
 105. Müller MJ, Geisler C, Hübers M, Pourhassan M, Braun W, Bosy-Westphal A. Normalizing resting energy expenditure across the life course in humans: Challenges and hopes. *Eur J Clin Nutr.* 2018;72(5):628–37.
 106. Cameron VA, Autelitano DJ, Evans JJ, Ellmers LJ, Espiner EA, Gary Nicholls M, et al. Body-size dependence of resting energy expenditure can be attributed to nonenergetic homogeneity of fat-free mass. *Am J Physiol - Endocrinol Metab.* 2002;282(1):E132-8.
 107. Kaplowitz PB, Slora EJ, Wasserman RC, Pedlow SE, Herman-Giddens ME. Earlier onset of puberty in girls: Relation to increased body mass index and race. *Pediatrics.* 2001;108(2):347–53.
 108. Lee SJ, Arslanian SA. Fat oxidation in black and white youth: A metabolic phenotype potentially predisposing black girls to obesity. *J Clin Endocrinol Metab.* 2008;93(11):4547–51.
 109. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr.* 1985;39C(Suppl 1):5–41.

110. Schofield WN, Schofield C, James WPT. Basal metabolic rate - Review and prediction, together with an annotated bibliography of source material. *Hum Nutr Clin Nutr.* 1985;39C(Suppl 1):1-96.
111. Lamonte MJ, Ainsworth BE. Quantifying energy expenditure and physical activity in the context of dose response. *Med Sci Sports Exerc.* 2001;33(Suppl 6):S370-8; discussion S419-20.
112. Thorburn AW, Proietto J. Biological determinants of spontaneous physical activity. *Obes Rev.* 2000;1(2):87-94.
113. Leibel RL, Rosenbaum M, Hirsch J. Changes in Energy Expenditure Resulting from Altered Body Weight. *N Engl J Med.* 1995;332(10):621-8.
114. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc.* 2000;32(Suppl 9):S498-504.
115. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: A second update of codes and MET values. *Med Sci Sports Exerc.* 2011;43(8):1575-81.
116. Ainsworth BE, Watson KB, Ridley K, Pfeiffer KA, Herrmann SD, Crouter SE, et al. Utility of the Youth Compendium of Physical Activities. *Res Q Exerc Sport.* 2018;89(3):273-81.
117. Butte NF, Watson KB, Ridley K, Zakeri IF, McMurray RG, Pfeiffer KA, et al. A youth compendium of physical activities: Activity codes and metabolic intensities. *Med Sci Sports Exerc.* 2018;50(2):246-56.
118. Ridley K, Olds TS. Assigning energy costs to activities in children: A review and synthesis. *Med Sci Sports Exerc.* 2008;40(8):1439-46.
119. Pfeiffer KA, Watson KB, McMurray RG, Bassett DR, Butte NF, Crouter SE,

- et al. Energy cost expression for a youth compendium of physical activities: Rationale for using age groups. *Pediatr Exerc Sci*. 2018;30(1):144–51.
120. Ridley K, Ainsworth BE, Olds TS. Development of a Compendium of Energy Expenditures for youth. *Int J Behav Nutr Phys Act*. 2008;5(1):45.
 121. Herrmann SD, Pfeiffer KA. New data for an updated Youth Energy Expenditure Compendium: an introduction. *J Phys Act Heal*. 2016;13(Suppl 1):S1–2.
 122. Trost SG, Drovandi CC, Pfeiffer K. Developmental trends in the energy cost of physical activities performed by youth. *J Phys Act Heal*. 2016;13(Suppl 1):S35–40.
 123. Youth Energy Expenditure Workshop [Internet]. Atlanta, GA: National Collaboration on Childhood Obesity Research (NCCOR); Available from: http://www.nccor.org/downloads/YEE_full_summary_FINAL.pdf
 124. McMurray RG, Butte NF, Crouter SE, Trost SG, Pfeiffer KA, Bassett DR, et al. Exploring Metrics to Express Energy Expenditure of Physical Activity in Youth. *PLoS One*. 2015;10(6):e0130869.
 125. Remmers T, Sleddens EFC, Gubbels JS, De Vries SI, Mommers M, Penders J, et al. Relationship between physical activity and the development of body mass index in children. *Med Sci Sports Exerc*. 2014;46(1):177–84.
 126. Jago R, Baranowski T, Baranowski J, Thompson D, Greaves K. BMI from 3-6 y of age is predicted by TV viewing and physical activity, not diet. *Int J Obes*. 2005;29(6):557–64.
 127. Moore L, Gao D, Bradlee M, Cupples L, Sundarajan-Ramamurti A, Proctor M, et al. Does early physical activity predict body fat change throughout childhood? *Prev Med*. 2003;37(1):10–7.

128. Jackowski SA, Baxter-Jones ADG, Gruodyte-Raciene R, Kontulainen SA, Erlandson MC. A longitudinal study of bone area, content, density, and strength development at the radius and tibia in children 4–12 years of age exposed to recreational gymnastics. *Osteoporos Int.* 2015;26(6):1677–90.
129. Janz KF, Letuchy EM, Eichenberger Gilmore JM, Burns TL, Torner JC, Willing MC, et al. Early physical activity provides sustained bone health benefits later in childhood. *Med Sci Sports Exerc.* 2010;42(6):1072–8.
130. Janz KF, Gilmore JME, Levy SM, Letuchy EM, Burns TL, Beck TJ. Physical activity and femoral neck bone strength during childhood: The Iowa Bone Development Study. *Bone.* 2007;41(2):216–22.
131. DuRant RH, Baranowski T, Rhodes T, Gutin B, Thompson WO, Carroll R, et al. Association among serum lipid and lipoprotein concentrations and physical activity, physical fitness, and body composition in young children. *J Pediatr.* 1993;123(2):185–92.
132. Knowles G, Pallan M, Thomas GN, Ekelund U, Cheng KK, Barrett T, et al. Physical activity and blood pressure in primary school children: A longitudinal study. *Hypertension.* 2013;61(1):70–5.
133. Beets MW, Beighle A, Erwin HE, Huberty JL. After-School Program Impact on Physical Activity and Fitness. A Meta-Analysis. *Am J Prev Med.* 2009;36(6):527–37.
134. Saavedra JM, Escalante Y, Garcia-Hermoso A. Improvement of aerobic fitness in obese children: A meta-analysis. *Int J Pediatr Obes.* 2011;6(3–4):169–77.
135. Dobbins M, Husson H, Decorby K, Larocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. *Cochrane Database Syst Rev.*

2013;28(2):CD007651.

136. Clark J. Does the type of intervention method really matter for combating childhood obesity? A systematic review and meta-analysis. *J Sport Med Phys Fit.* 2015;55(12):1524–1543.
137. Malina RM. Weight training in youth-growth, maturation, and safety: An evidence-based review. *Clin J Sport Med.* 2006;16(6):478–87.
138. Vasconcellos F, Seabra A, Katzmarzyk PT, Kraemer-Aguiar LG, Bouskela E, Farinatti P. Physical activity in overweight and obese adolescents: Systematic review of the effects on physical fitness components and cardiovascular risk factors. *Sport Med.* 2014;44(8):1139–52.
139. Millard-Stafford M, Becasen JS, Beets MW, Nihiser AJ, Lee SM, Fulton JE. Is Physical Fitness Associated With Health in Overweight and Obese Youth? A Systematic Review. *Kinesiol Rev.* 2016;2(4):233–47.
140. Faigenbaum AD, McFarland JE, Johnson L, Kang J, Bloom J, Ratamess NA, et al. Preliminary evaluation of an after-school resistance training program for improving physical fitness in middle school-age boys. *Percept Mot Skills.* 2007;104(2):407–15.
141. Wilks DC, Sharp SJ, Ekelund U, Thompson SG, Mander AP, Turner RM, et al. Objectively measured physical activity and fat mass in children: A bias-adjusted meta-analysis of prospective studies. *PLoS One.* 2011;6(2):e17205.
142. Nogueira RC, Weeks BK, Beck BR. Exercise to improve pediatric bone and fat: A systematic review and meta-analysis. *Med Sci Sports Exerc.* 2014;46(3):610–21.
143. Ramires VV, Dumith SC, Gonçalves H. Longitudinal association between physical activity and body fat during adolescence: A systematic review.

- J Phys Act Heal. 2015;12(9):1344–58.
144. Janssen I, LeBlanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int J Behav Nutr Phys Act.* 2010;7(1):40.
 145. Tan VPS, Macdonald HM, Kim SJ, Nettlefold L, Gabel L, Ashe MC, et al. Influence of physical activity on bone strength in children and adolescents: A systematic review and narrative synthesis. *J Bone Miner Res.* 2014;29(10):2161–81.
 146. Weaver CM, Gordon CM, Janz KF, Kalkwarf HJ, Lappe JM, Lewis R, et al. The National Osteoporosis Foundation’s position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporos Int.* 2016;27(4):1281–386.
 147. Kelley GA, Kelley KS. Aerobic exercise and lipids and lipoproteins in children and adolescents: A meta-analysis of randomized controlled trials. *Atherosclerosis.* 2007;191(2):447–53.
 148. Fedewa M V., Gist NH, Evans EM, Dishman RK. Exercise and insulin resistance in youth: A meta-analysis. *Pediatrics.* 2014;133(1):e163-74.
 149. Ahn S, Fedewa AL. A meta-analysis of the relationship between children’s physical activity and mental health. *J Pediatr Psychol.* 2011;36(4):385–97.
 150. Biddle SJH, Asare M. Physical activity and mental health in children and adolescents: a review of reviews. *Br J Sports Med.* 2011;45(11):886–95.
 151. Pastor Y, Balaguer I, Pons D, García-Merita M. Testing direct and indirect effects of sports participation on perceived health in Spanish adolescents between 15 and 18 years of age. *J Adolesc.* 2003;26(6):717–

30.

152. Parfitt G, Eston RG. The relationship between children's habitual activity level and psychological well-being. *Acta Paediatr.* 2007;94(12):1791-7.
153. Ruiz-Hermosa A, Álvarez-Bueno C, Cavero-Redondo I, Martínez-Vizcaíno V, Redondo-Tébar A, Sánchez-López M. Active commuting to and from school, cognitive performance, and academic achievement in children and adolescents: A systematic review and meta-analysis of observational studies. *Int J Environ Res Public Health.* 2019;16(10):1839.
154. Álvarez-Bueno C, Pesce C, Cavero-Redondo I, Sánchez-López M, Martínez-Hortelano JA, Martínez-Vizcaíno V. The Effect of Physical Activity Interventions on Children's Cognition and Metacognition: A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry.* 2017;56(9):729-38.
155. Sánchez-López M, Cavero-Redondo I, Álvarez-Bueno C, Ruiz-Hermosa A, Pozuelo-Carrascosa DP, Díez-Fernández A, et al. Impact of a multicomponent physical activity intervention on cognitive performance: The MOVI-KIDS study. *Scand J Med Sci Sport.* 2019;29(5):766-75.
156. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput JP, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: An update. *Appl Physiol Nutr Metab.* 2016;41(6):S240-65.
157. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. *Int J Behav Nutr Phys Act.* 2011;8:98.

158. Cliff DP, Hesketh KD, Vella SA, Hinkley T, Tsiros MD, Ridgers ND, et al. Objectively measured sedentary behaviour and health and development in children and adolescents: Systematic review and meta-analysis. *Obes Rev.* 2016;17(4):330–44.
159. Heidemann M, Mølgaard C, Husby S, Schou AJ, Klakk H, Møller NC, et al. The intensity of physical activity influences bone mineral accrual in childhood: The childhood health, activity and motor performance school (the CHAMPS) study, Denmark. *BMC Pediatr.* 2013;13(1):32.
160. Gabel L, Macdonald HM, Nettlefold L, McKay HA. Physical Activity, Sedentary Time, and Bone Strength From Childhood to Early Adulthood: A Mixed Longitudinal HR-pQCT study. *J Bone Miner Res.* 2017;32(7):1525–36.
161. Schoeller DA, Van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol Respir Environ Exerc Physiol.* 1982;53(4):955–9.
162. Schoeller DA, Webb P. Five-day comparison of the doubly labeled water method with respiratory gas exchange. *Am J Clin Nutr.* 1984;40(1):153–8.
163. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jéquier E. Energy expenditure by doubly labeled water: Validation in humans and proposed calculation. *Am J Physiol - Regul Integr Comp Physiol.* 1986;250(5 Pt 2):R823-30.
164. Jones PJH, Winthrop AL, Schoeller DA, Swyer PR, Smith J, Filler RM, et al. Validation of doubly labeled water for assessing energy expenditure in Infants. *Pediatr Res.* 1987;21(3):242–6.
165. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity.*

- 2007;15(10):2371-9.
166. Speakman JR. The history and theory of the doubly labeled water technique. *Am J Clin Nutr.* 1998;68(4):S932-8.
 167. Chen LF, Vander Weg MW, Hofmann DA, Reisinger HS. The Hawthorne effect in infection prevention and epidemiology. *Infect Control Hosp Epidemiol.* 2015;36(12):1444-50.
 168. Laporte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Rep.* 1985;100(2):131-46.
 169. Da Rocha EEM, Alves VGF, Da Fonseca RB V. Indirect calorimetry: Methodology, instruments and clinical application. *Curr Opin Clin Nutr Metab Care.* 2006;9(3):247-56.
 170. Haugen AH, Chan LN, Li F. Indirect calorimetry: A practical guide for clinicians. *Nutr Clin Pract.* 2007;22(4):377-88.
 171. Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, et al. How to assess physical activity? How to assess physical fitness? *Eur J Prev Cardiol.* 2005;12(2):102-14.
 172. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil.* 2010;17(2):127-39.
 173. Bassett DR. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport.* 2000;71(Suppl 2):S30-36.
 174. Ainsworth B, Cahalin L, Buman M, Ross R. The Current State of Physical

- Activity Assessment Tools. *Prog Cardiovasc Dis.* 2015;57(4):387–95.
175. Cicone ZS, Holmes CJ, Fedewa M V., MacDonald H V., Esco MR. Age-Based Prediction of Maximal Heart Rate in Children and Adolescents: A Systematic Review and Meta-Analysis. *Res Q Exerc Sport.* 2019;90(3):417–28.
 176. Healey J. Future possibilities in electronic monitoring of physical activity. *Res Q Exerc Sport.* 2000;71(Suppl 2):S137-45.
 177. Pols MA, Peeters PHM, Kemper HCG, Grobbee DE. Methodological aspects of physical activity assessment in epidemiological studies. *Eur J Epidemiol.* 1998;14(1):63–70.
 178. Ainsworth BE. How do i measure physical activity in my patients? Questionnaires and objective methods. *Br J Sports Med.* 2009;43(1):6–9.
 179. Melanson EL, Knoll JR, Bell ML, Donahoo WT, Hill JO, Nysse LJ, et al. Commercially available pedometers: Considerations for accurate step counting. *Prev Med (Baltim).* 2004;39(2):361–8.
 180. Crouter SE, Schneider PL, Karabulut M, Bassett DR. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med Sci Sports Exerc.* 2003;35(8):1455–60.
 181. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: Convergent validity. *Sport Med.* 2002;32(12):795–808.
 182. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: Construct validity. *Sport Med.* 2004;34(5):281–91.
 183. Chen KY, Bassett DR. The Technology of Accelerometry-Based Activity

- Monitors: Current and Future. *Med Sci Sport Exerc.* 2005;37(Suppl 11):S490-500.
184. Borghese MM, Borgundvaag E, McIsaac MA, Janssen I. Imputing accelerometer nonwear time in children influences estimates of sedentary time and its associations with cardiometabolic risk. *Int J Behav Nutr Phys Act.* 2019;16(1):7.
 185. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc.* 2011;43(7):1360–8.
 186. Lagerros YT. Physical activity-the more we measure, the more we know how to measure. *Eur J Epidemiol.* 2009;24(3):119–22.
 187. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: Lessons from cognitive interviews. *Med Sci Sports Exerc.* 2009;41(2):336–43.
 188. Baranowski T, Domel SB. A cognitive model of children’s reporting of food intake. *Am J Clin Nutr.* 1994;59(Suppl 1):S212-217.
 189. Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. *Appl Cogn Psychol.* 1991;5(3):251–67.
 190. Patterson P. Reliability, validity, and methodological response to the assessment of physical activity via self-report. *Res Q Exerc Sport.* 2000;71(Suppl 2):S15-20.
 191. Kesaniemi YA, Danforth E. J, Jensen MD, Kopelman PG, Lefebvre P, Reeder BA. Dose-response issues concerning physical activity and health: An evidence-based symposium. *Med Sci Sports Exerc.* 2001;33(Suppl 6):S351-8.

192. Terwee CB, Mokkink LB, Van Poppel MNM, Chinapaw MJM, Van Mechelen W, De Vet HCW. Qualitative Attributes and Measurement Properties of Physical Activity Questionnaires A Checklist. *Sport Med.* 2010;40(7):525–37.
193. Sternfeld B, Goldman-Rosas L. A systematic approach to selecting an appropriate measure of self-reported physical activity or sedentary behavior. *J Phys Act Health.* 2012;9(Suppl 1):S19-28.
194. Forsén L, Loland NW, Vuillemin A, Chinapaw MJM, Van Poppel MNM, Mokkink LB, et al. Self-administered physical activity questionnaires for the elderly: A systematic review of measurement properties. *Sport Med.* 2010;40(7):601–23.
195. Van Poppel MNM, Chinapaw MJM, Mokkink LB, Van Mechelen W, Terwee CB. Physical activity questionnaires for adults: A systematic review of measurement properties. *Sport Med.* 2010;40(7):565–600.
196. Bowles HR. Measurement of active and sedentary behaviors: closing the gaps in self-report methods. *J Phys Act Health.* 2012;9(Suppl 1):S1–4.
197. Coleman KJ, Ngor E, Reynolds K, Quinn VP, Koebnick C, Young DR, et al. Initial validation of an exercise “vital sign” in electronic medical records. *Med Sci Sports Exerc.* 2012;44(11):2071–6.
198. Physical Activity Resource Center for Public Health [Internet]. University of Pittsburgh. Web Systems by Epidemiology Data Center. Available from: <http://www.parcph.org/>
199. Hagströmer M, Ainsworth BE, Kwak L, Bowles HR. A checklist for evaluating the methodological quality of validation studies on self-report instruments for physical activity and sedentary behavior. *J Phys Act Health.* 2012;9(Suppl 1):S29–36.

200. Ainsworth BE, Caspersen CJ, Matthews CE, Mâsse LC, Baranowski T, Zhu W. Recommendations to improve the accuracy of estimates of physical activity derived from self report. *J Phys Act Health*. 2012;9(Suppl 1):S76–S84.
201. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: practical guides to biostatistics and epidemiology*. Cambridge University Press; 2011.
202. Rennie KL, Wareham NJ. The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. *Public Health Nutr*. 1998;1(4):265–71.
203. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–95.
204. Nishikido N, Kashiwazaki H, Suzuki T. Preschool children's daily activities: direct observation, pedometry or questionnaire. *J Hum Ergol (Tokyo)*. 1982;11(2):214–8.
205. Telford A, Salmon J, Jolley D, Crawford D. Reliability and Validity of Physical Activity Questionnaires for Children: The Children's Leisure Activities Study Survey (CLASS). *Pediatr Exerc Sci*. 2004;16(1):64–78.
206. Koo MM, Rohan TE. Comparison of four habitual physical activity questionnaires in girls aged 7-15 yr. *Med Sci Sports Exerc*. 1999;31(3):421–7.
207. Janz KF, Witt J, Mahoney LT. The stability of children's physical activity as measured by accelerometry and self-report. *Med Sci Sports Exerc*. 1995;27(9):1326–32.
208. Kowalski KC, Crocker PRE, Faulkner RA. Validation of the Physical

- Activity Questionnaire for Older Children. *Pediatr Exerc Sci*. 1997;9(2):174–86.
209. Sallis JF, Strikmiller PK, Harsha DW, Feldman HA, Ehlinger S, Stone EJ, et al. Validation of interviewer- and self-administered physical activity checklists for fifth grade students. *Med Sci Sports Exerc*. 1996;28(7):840–51.
210. Harro M. Validation of a questionnaire to assess physical activity of children ages 4–8 years. *Res Q Exerc Sport*. 1997;68(4):259–68.
211. Chen X, Sekine M, Hamanishi S, Wang H, Gaina A, Yamagami T, et al. Validation of a self-reported physical activity questionnaire for schoolchildren. *J Epidemiol*. 2003;13(5):278–87.
212. Bere E, Bjørkelund LA. Test-retest reliability of a new self reported comprehensive questionnaire measuring frequencies of different modes of adolescents commuting to school and their parents commuting to work - The ATN questionnaire. *Int J Behav Nutr Phys Act*. 2009;6:68.
213. Shephard RJ. Absolute versus relative intensity of physical activity in a dose-response context. *Med Sci Sports Exerc*. 2001;33(Suppl 6):S400-420.
214. Kremers SPJ, Visscher TLS, Seidell JC, Van Mechelen W, Brug J. Cognitive determinants of energy balance-related behaviours: Measurement issues. *Sport Med*. 2005;35(11):923–33.
215. Janz KF, Broffitt B, Levy SM. Validation evidence for the netherlands physical activity questionnaire for young children: The Iowa bone development study. *Res Q Exerc Sport*. 2005;76(3):363–9.
216. Treuth MS, Hou N, Young DR, Maynard LM. Validity and reliability of the

- Fels physical activity questionnaire for children. *Med Sci Sports Exerc.* 2005;37(3):488–95.
217. Chasan-Taber L, Erickson J, McBride J, Nasca P, Chasan-Taber S, Freedson P. Reproducibility of a self-administered lifetime physical activity questionnaire among female college alumnae. *Am J Epidemiol.* 2002;155(3):282–9.
218. Rifas-Shiman SL, Gillman MW, Field AE, Frazier AL, Berkey CS, Tomeo CA, et al. Comparing physical activity questionnaires for youth. Seasonal vs annual format. *Am J Prev Med.* 2001;20(4):282–5.
219. Corder K, Van Sluijs EMF, Wright A, Whincup P, Wareham NJ, Ekelund U. Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? *Am J Clin Nutr.* 2009;89(3):862–70.
220. Burdette HL, Whitaker RC, Daniels SR. Parental Report of Outdoor Playtime as a Measure of Physical Activity in Preschool-aged Children. *Arch Pediatr Adolesc Med.* 2004;158(4):353–7.
221. Moore HJ, Ells LJ, Mclure SA, Crooks S, Cumbor D, Summerbell CD, et al. The development and evaluation of a novel computer program to assess previous-day dietary and physical activity behaviours in school children: The Synchronised Nutrition and Activity Program™ (SNAP™). *Br J Nutr.* 2008;99(6):1266–74.
222. Nusser SM, Beyler NK, Welk GJ, Carriquiry AL, Fuller WA, King B MN. Modeling errors in physical activity recall data. *J Phys Act Health.* 2012;9(Suppl 1):S56–67.
223. Verheul ACM, Prins AN, Kemper HCG, Kardinaal AFM, Van Erp-Baart MAJ. Validation of a weight-bearing physical activity questionnaire in a study of bone density in girls and women. *Pediatr Exercise Sci.*

- 1998;10(1):38-47.
224. Bingham DD, Collings PJ, Clemes SA, Costa S, Santorelli G, Griffiths P, et al. Reliability and Validity of the Early Years Physical Activity Questionnaire (EY-PAQ). *Sport* (Basel, Switzerland). 2016 May;4(2):30.
225. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38(1):27-36.
226. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther*. 1996;18(5):979-92.
227. Fayers P, Hand D. Factor analysis, causal indicators and quality of life. *Qual Life Res*. 1997;6:139-50.
228. Edwards J, Bagozzi R. On the nature and direction of relationships between constructs and measures. *Psychol Rev*. 2000;5:155-74.
229. Buch-Andersen T, Pérez-Cueto FJA, Toft U. Relative validity and reproducibility of a parent-administered semi-quantitative FFQ for assessing food intake in Danish children aged 3-9 years. *Public Health Nutr*. 2016;19(7):1184-94.
230. Podnar H, Sušilović MK, Radman I. Internal consistency and retest reliability of the Croatian version of PAQ-C. *Croat J Educ*. 2017;19(1):45-62.
231. Hardy LL, Booth ML, Okely AD. The reliability of the Adolescent Sedentary Activity Questionnaire (ASAQ). *Prev Med (Baltim)*. 2007;45(1):71-4.
232. Storey KE, Mccargar LJ. Reliability and validity of Web-SPAN, a web-based method for assessing weight status, diet and physical activity in

- youth. *J Hum Nutr Diet.* 2012;25(1):59–68.
233. Liu Y, Wang M, Tynjälä J, Lv Y, Villberg J, Zhang Z, et al. Test-retest reliability of selected items of Health Behaviour in School-aged Children (HBSC) survey questionnaire in Beijing, China. *BMC Med Res Methodol.* 2010;10(10):73.
234. Stanley RM, Ridley K, Olds TS, Dollman J. Development and psychometric properties of the Y-PASS questionnaire to assess correlates of lunchtime and after-school physical activity in children. *BMC Public Health.* 2014;14(1):412.
235. Williams K, Frei A, Vetsch A, Dobbels F, Puhan MA, Rüdell K. Patient-reported physical activity questionnaires: A systematic review of content and format. *Health Qual Life Outcomes.* 2012;10:28.
236. Isa T, Sawa R, Torizawa K, Murata S, Saito T, Ebina A, et al. Reliability and Validity of the Japanese Version of the Physical Activity Questionnaire for Older Children. *Clin Med Insights Pediatr.* 2019;13:1–6.
237. Singh AS, Vik FN, Chinapaw MJ, Uijtdewilligen L, Verloigne M, Fernandez-Alvira JM, et al. Test-retest reliability and construct validity of the ENERGY-child questionnaire on energy balance-related behaviours and their potential determinants: the ENERGY-project. *Int J Behav Nutr Phys Act.* 2011;8:136.
238. Mokkink LB, Prinsen CACC, Bouter LM, de Vet H.C.W, Terwee CB, de Vet HCW, et al. The Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian J Phys Ther.* 2016;20(2):105–13.
239. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy,

- terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010 Jul 1;63(7):737–45.
240. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res.* 2002;11(3):193–205.
241. American Psychological Association. National Council on Measurement in Education. Standards for Educational and Psychological Testing [Internet]. 1999. Available from: <https://www.apa.org/science/programs/testing/standards>
242. 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 Jan 1;60(1):34–42.
243. Terwee CB, Prinsen CA, Chiarotto A, Cw De Vet H, Bouter LM, Marjan JA, et al. COSMIN standards and criteria for evaluating the content validity of health-related Patient-Reported Outcome Measures: a Delphi study. *Qual Life Res Press.*
244. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159–70.
245. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Heal.* 2009;12(8):1075–

83.

246. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995 Aug;4(4):293–307.
247. 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(5):1147–57.
248. Hidding LM, Altenburg TM, Mokkink LB, Terwee CB, Chinapaw MJM. Systematic Review of Childhood Sedentary Behavior Questionnaires: What do We Know and What is Next? *Sport Med.* 2017;47(4):677–99.
249. Arredondo EM, Mendelson T, Holub C, Espinoza N, Marshall S. Cultural adaptation of physical activity self-report instruments. *J Phys Act Health.* 2012;9(Suppl 1):S37–S43.
250. 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.
251. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59(10):1033–9.
252. Sallis JF. Self-Report Measures of Children’s Physical Activity. *J Sch Health.* 1991;61(5):215–9.
253. Crocker PRE, Bailey DA, Faulkner RA, Kowalski KC, Mcgrath R. Measuring general levels of physical activity: Preliminary evidence for the physical activity questionnaire for older children. *Med Sci Sports Exerc.* 1997;29(10):1344–9.
254. Tremblay MS, Wyatt Inman J, Douglas Willms J. Preliminary evaluation of a video questionnaire to assess activity levels of children. *Med Sci*

- Sports Exerc. 2001;33(12):2139–44.
255. Treuth MS, Sherwood NE, Butte NF, McClanahan B, Obarzanek E, Zhou A, et al. Validity and reliability of activity measures in African-American Girls for GEMS. *Med Sci Sports Exerc.* 2003;35(3):532–9.
256. Simons-Morton BG, Taylor WC, Huang IW. Validity of the physical activity interview and caltrac with preadolescent children. *Res Q Exerc Sport.* 1994;65(1):84–8.
257. McMurray RG, Harrell JS, Bradley CB, Webb JP, Goodman EM. Comparison of a computerized physical activity recall with a triaxial motion sensor in middle-school youth. *Med Sci Sports Exerc.* 1998;30(8):1238–45.
258. Ridley K, Olds TS, Hill A. The Multimedia activity recall for children and adolescents (MARCA): Development and evaluation. *Int J Behav Nutr Phys Act.* 2006;3:10.
259. Sánchez P, Hernández B. Validación de un cuestionario de actividad física habitual. *Apunt Educ física y Deport.* 1999;58:68–72.
260. Eisenmann J, Milburn N, Jacobsen L, Moore SJ. Reliability and convergent validity of the Godin Leisure-Time Exercise Questionnaire in rural 5th-grade school-children. *J Hum Mov Stud.* 2002;43:135–49.
261. González-Gil EM, Mouratidou T, Cardon G, Androustos O, De Bourdeaudhuij I, Gózdź M, et al. Reliability of primary caregivers reports on lifestyle behaviours of European pre-school children: The ToyBox-study. *Obes Rev.* 2014;15(Suppl 3):61–6.
262. Wang JJ, Baranowski T, Lau WCP, Chen TA, Pitkethly AJ. Validation of the Physical Activity Questionnaire for Older Children (PAQ-C) among Chinese Children. *Biomed Environ Sci.* 2016;29(3):177–86.

263. Yang X, Zhai Y, Si X, Zhao WH. Validity and reliability of physical activity questionnaires in children and adolescents: a Meta-analysis. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2020;54(5):546–54.
264. International Physical Activity Questionnaire [Internet]. Available from: <https://sites.google.com/site/theipaq/>
265. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Vol. 6, *PLoS Medicine*. 2009.
266. Search blocks. Biomedische Informatie. Real Asociación Holandesa de Profesionales de la Información (KNVI) [Internet]. [cited 2020 Oct 9]. Available from: <https://blocks.bmi-online.nl/>
267. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009 Oct 27;18(8):1115–23.
268. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–58.
269. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J*. 2003;327(7414):557–60.
270. Moore JB, Hanes JC, Barbeau P, Gutin B, Treviño RP, Yin Z. Validation of the physical activity questionnaire for older children in children of different races. *Pediatr Exerc Sci*. 2007;19(1):6–19.
271. Faghihimani Z, Nourian M, Nikkar AH, Farajzadegan Z, Khavariyan N, Ghatrehsamani S, et al. Validation of the child and adolescent international physical activity questionnaires in iranian children and adolescents. *ARYA Atheroscler*. 2010;5(4):163–6.

272. Herazo-Beltrán AY, Domínguez-Anaya R. The reliability of a questionnaire regarding Colombian children's physical activity. *Rev salud pública*. 2012;14(5):802–9.
273. Benítez-Porres J, López-Fernández I, Raya JF, Álvarez Carnero S, Alvero-Cruz JR, Álvarez Carnero E. Reliability and Validity of the PAQ-C Questionnaire to Assess Physical Activity in Children. *J Sch Health*. 2016 Sep 1;86(9):677–85.
274. Manchola-González J, Bagur-Calafat C, Girabent-Farrés M. Reliability of the Spanish Version of Questionnaire of Physical Activity PAQ-C. *Rev Int Med Y Ciencias La Act Física Y Del Deport*. 2017;17(65):139–52.
275. Vidaković Samaržija D, M. M-D. Reliability of croatian version of the questionnaire for assessment of overall level of physical activity of younger school children. *Croat Sport Med J*. 2013;28(1):24–32.
276. Thomas EL, Upton D. Psychometric properties of the physical activity questionnaire for older children (PAQ-C) in the UK. *Psychol Sport Exerc*. 2014 May 1;15(3):280–7.
277. Bervoets L, Van Noten C, Van Roosbroeck S, Hansen D, Van Hoorenbeeck K, Verheyen E, et al. Reliability and Validity of the Dutch Physical Activity Questionnaires for Children (PAQ-C) and Adolescents (PAQ-A). *Arch Public Heal*. 2014;72:47.
278. Guedes DP, Guedes JERP. Medida da atividade física em jovens brasileiros: Reprodutibilidade e validade do PAQ-C e do PAQ-A. *Rev Bras Med do Esporte*. 2015;21(6):425–32.
279. Zaki NAM, Sahril N, Omar MA, Ahmad MH, Baharudin A, Nor NSM. Reliability and validity of the physical activity questionnaire for older children (PAQ-C) in Malay language. *Int J Public Heal Res*. 2015;6(1):670–6.

280. Gobbi E, Elliot C, Varnier M, Carraro A. Psychometric Properties of the Physical Activity Questionnaire for Older Children in Italy: Testing the Validity among a General and Clinical Pediatric Population. *PLoS One*. 2016;11(5):e0156354.
281. Malan GF, Nolte K. Measuring physical activity in south African grade 2 and 3 learners: a self-report questionnaire versus pedometer testing. *South African J Res Sport Phys Educ Recreat*. 2017;39(3):79–91.
282. Ben Jemaa H, Mankaï A, Mahjoub F, Kortobi B, Khlifi S, Draoui J, et al. Physical Activity Level Assessed by Accelerometer and PAQ-C in Tunisian Children. *Ann Nutr Metab*. 2018;73(3):234–40.
283. Erdim L, Ergün A, Kuğuoğlu S. Reliability and validity of the turkish version of the physical activity questionnaire for older children (Paq-C). *Turkish J Med Sci*. 2019;49(1):162–9.
284. Zishan KB, Sheth M, Sharma S. Reliability and validity of Gujarati version of physical activity questionnaire for children. *Int J Phys Educ Sport Heal*. 2019;6(4):154–7.
285. Pratanaphon S, Longlalerng K, Jitmana J, Chaikla K, Nankanya T, Pirunsarn U, et al. Content validity and psychometric characteristics of the Thai translated version of the physical activity questionnaire for children (PAQ-C) and adolescents (PAQ-A). *J Assoc Med Sci*. 2020;53(3):84–91.
286. Hnatiuk JA, Salmon J, Hinkley T, Okely AD, Trost S. A review of preschool children's physical activity and sedentary time using objective measures. *Am J Prev Med*. 2014;47(4):487–97.
287. Martín-Bello C, Vicente-Rodriguez G, Casajús-Mallén J, Gómez-Bruton A. Validación de los cuestionarios PAQ-C e IPAQ-A en niños/as en edad escolar. *Cult Cienc y Deport*. 2020;15(44):177–87.

288. Berrigan D, Forsyth BH, Helba C, Levin K, Norberg A, Willis GB. Cognitive testing of physical activity and acculturation questions in recent and long-term Latino immigrants. *BMC Public Health*. 2010;10:481.
289. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186-91.
290. Martínez-Gómez D, Martínez-De-Haro V, Pozo T, Welk GJ, Villagra A, Calle ME, et al. Reliability and validity of the PAQ-A questionnaire to assess physical activity in Spanish adolescents. *Rev Esp Salud Publica*. 2009;83(3):427-39.
291. Knauper B, Belli R, Hill D, Herzog A. Question Difficulty and Respondents' Cognitive Ability: The Impact on Data Quality. *J Off Stat*. 1997;13(2):181-99.
292. Janz KF, Lutuchy EM, Wenthe P, Levy SM. Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. *Med Sci Sports Exerc*. 2008;40(4):767-72.
293. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *J Eval Clin Pract*. 2011;17(2):268-74.
294. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res*. 2010 May;19(4):539-49.
295. Dwyer GM, Hardy LL, Peat JK, Baur LA. The validity and reliability of a home environment preschool-age physical activity questionnaire (Pre-PAQ). *Int J Behav Nutr Phys Act*. 2011;8:86.

296. Bonn SE, Surkan PJ, Trolle Lagerros Y, Bälter K. Feasibility of a novel web-based physical activity questionnaire for young children. *Pediatr Rep.* 2012;4(4):e37.
297. Nascimento-Ferreira MV, De Moraes ACF, Toazza-Oliveira PV, Forjaz CLM, Aristizabal JC, Santaliesra-Pasías AM, et al. Reliability and Validity of a Questionnaire for Physical Activity Assessment in South American Children and Adolescents: The SAYCARE Study. *Obesity.* 2018;26:S23–30.
298. Zelener J, Schneider M. Adolescents and Self-Reported Physical Activity: An Evaluation of the Modified Godin Leisure-Time Exercise Questionnaire. *Int J Exerc Sci.* 2016;9(5):587–98.
299. Huang YJ, Wong SHS, Salmon J. Reliability and validity of the modified Chinese version of the Children’s Leisure Activities Study Survey (CLASS) questionnaire in assessing physical activity among Hong Kong children. *Pediatr Exerc Sci.* 2009;21(3):339–53.
300. Sallis J, Saelens B. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport.* 2000;71(Suppl 2):S1-14.
301. Saris WHM. The assessment and evaluation of daily physical activity in children. A review. *Acta Pædiatrica.* 1985;74:37–48.
302. Costa S, Barber SE, Cameron N, Clemes SA. Calibration and validation of the ActiGraph GT3X+ in 2-3 year olds. *J Sci Med Sport.* 2014;17(6):617–22.
303. Rice KR, Joschtel B, Trost SG. Validity of family child care providers’ proxy reports on children’s physical activity. *Child Obes.* 2013 Oct 1;9(5):393–8.

304. Sallis JF, Patterson TL, Buono MJ, Nader PR. Relation of cardiovascular fitness and physical activity to cardiovascular disease risk factors in children and adults. *Am J Epidemiol.* 1988;127(5):933–41.
305. Saris W, Binkhorst A, Cramwinckel F, Van-Waesberghe A, Vam-der-Veen-Hezemans A. The relationship between working performance, daily physical activity, fatness, blood lipids, and nutrition in schoolchildren. . In: *Chil.* 1980.
306. Sallis JF, Buono MJ, Roby JJ, Micale FG, Nelson JA. Seven-day recall and other physical activity self-reports in children and adolescents. *Med Sci Sports Exerc.* 1993;25(1):99–108.
307. Sallis JF, Haskell WL, Wood PD, Fortmann SP, Rogers T, Blair SN, et al. Physical activity assessment methodology in the five-city project. *Am J Epidemiol.* 1985;121(1):91–106.
308. Sallis JF, Buono MJ, Roby JJ, Carlson D, Nelson JA. The Caltrac accelerometer as a physical activity monitor for school-age children. *Med Sci Sports Exerc.* 1990;22(5):698–703.
309. Simons-Morton BG, O’Hara NM, Parcel GS, Wei Huang I, Baranowski T, Wilson B. Children’s frequency of participation in moderate to vigorous physical activities. *Res Q Exerc Sport.* 1990;61(4):307–14.
310. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci.* 1985;10(3):141–6.
311. Kremers SPJ, Brug J. Habit strength of physical activity and sedentary behavior among children and adolescents. *Pediatr Exerc Sci.* 2008;20(1):5–14.
312. Baranowski T, Dworkin RJ, Cieslik CJ, Hooks P, Clearman DR, Ray L, et al. Reliability and Validity of Self Report of Aerobic Activity: Family

- Health Project. *Res Q Exerc Sport*. 1984;55(4):309–17.
313. Thurn J, Finne E, Brandes M, Bucksch J. Validation of physical activity habit strength with subjective and objective criterion measures. *Psychol Sport Exerc*. 2014 Jan 1;15(1):65–71.
314. Kilanowski CK, Consalvi AR, Epstein LH. Validation of an electronic pedometer for measurement of physical activity in children. *Pediatr Exerc Sci*. 1999;11(1):63–8.
315. Phillips LRS, Parfitt G, Rowlands A V. Calibration of the GENEActiv accelerometer for assessment of physical activity intensity in children. *J Sci Med Sport*. 2013;16(2):124–8.
316. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res*. 2002;10(3):150–7.
317. Santos-Lozano A, Santín-Medeiros F, Cardon G, Torres-Luque G, Bailón R, Bergmeir C, et al. Actigraph GT3X: Validation and determination of physical activity intensity cut points. *Int J Sports Med*. 2013;34(11):975–82.
318. Sun DX, Schmidt G, Teo-Koh SM. Validation of the RT3 accelerometer for measuring physical activity of children in simulated free-living conditions. *Pediatr Exerc Sci*. 2008;20(2):181–97.
319. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14(5):411–6.
320. Kavouras S, Sarras S, Tsekouras Y, Sidossis L. Assessment of energy expenditure in children using the RT3 accelerometer. *J Sports Sci*. 2008;26(9):959–66.
321. Trost SG, Marshall AL, Miller R, Hurley JT, Hunt JA. Validation of a 24-h physical activity recall in indigenous and non-indigenous Australian

- adolescents. *J Sci Med Sport*. 2007;10(6):428–35.
322. Colley R, Wong S, Garriguet D, Janssen I, Connor-Gorber S, Tremblay M. Physical activity, sedentary behaviour and sleep in Canadian children: parent-report versus direct measures and relative associations with health risk. *Heal Rep* . 2012;23(2):45–52.
323. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. *J Sports Sci*. 2008;26(14):1557–65.
324. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. *Med Sci Sports Exerc*. 2005;37(Suppl 11):S523-30.
325. Schaefer CA, Nigg CR, Hill JO, Brink LA, Browning RC. Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth. *Med Sci Sports Exerc*. 2014;46(4):826–33.
326. Cabero-Almenara J, Infante Moro A. Empleo del método Delphi y su empleo en la investigación en comunicación y educación. *EduTec Rev electrónica Tecnol Educ*. 2014 Jun 26;0(48):3.
327. Reguant-Álvarez M, Torrado-Fonseca M. El método Delphi. *Revsta d’Innovació i Recer en Educ*. 2016;9(1):87–102.
328. Nunnaly J. *Psychometric Theory*. New York, McGraw-Hill; 1977.
329. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med*. 2012;31(29):3972–81.
330. Fleiss J, Levin B, Paik M C. *Determining Sample Sizes Needed to Detect a Difference between Two Proportions*. John Wiley & Sons, Ltd; 2004. 64–85 p.

331. ActiGraph wGT3X-BT | ActiGraph [Internet]. Available from: <https://actigraphcorp.com/actigraph-wgt3x-bt/>
332. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc.* 2011;43(2):357–64.
333. Instituto Nacional de Estadística (INE). Clasificación por nivel de estudios [Internet]. Available from: <https://www.ine.es/daco/daco42/discapa/tar3.pdf>
334. Girabent-Farrés M, Fagoaga J, Amor-Barbosa M, Bagur-Calafat C. Spanish translation and validation of the neuromuscular module of the pediatric quality of life inventory (PedsQL): Evaluation of the quality of life perceived by 8-18 years old children with neuromuscular diseases and by their parents. *Rev Neurol.* 2018;67(11):425–35.
335. Girabent-Farrés M, Bagur-Calafat C, Amor-Barbosa M, Natera-De Benito D, Medina-Rincón A, Fagoaga J. Spanish translation and validation of the Neuromuscular Module of the Pediatric Quality of Life Inventory (PedsQL): Evaluation of the quality of life perceived by 5-7 years old children with neuromuscular disorders and by their parents. *Rev Neurol.* 2019;69(11):442–52.
336. Leidy NK, Vernon M. Perspectives on patient-reported outcomes: Content validity and qualitative research in a changing clinical trial environment. *Pharmacoeconomics.* 2008;26(5):363–70.
337. Montoye H, Kemper H, Saris W, Washburn R. Measuring physical activity and energy expenditure. *Human Kinetics, Champaign, IL;* 1996.
338. Wong SL, Leatherdale ST, Manske S. Reliability and validity of a school-based physical activity questionnaire. *Med Sci Sports Exerc.* 2006;38(9):1593–600.

339. Bringolf-Isler B, Mäder U, Ruch N, Kriemler S, Grize L, Braun-Fahrländer C. Measuring and validating physical activity and sedentary behavior comparing a parental questionnaire to accelerometer data and diaries. *Pediatr Exerc Sci.* 2012;24(2):229–45.
340. Basterfield L, Adamson AJ, Parkinson KN, Maute U, Li PX, Reilly JJ, et al. Surveillance of physical activity in the UK is flawed: Validation of the Health Survey for England Physical Activity Questionnaire. *Arch Dis Child.* 2008;93(12):1054–8.
341. Kowalski C, Crocker P, Kowalski N. Convergent validity of the physical activity questionnaire for adolescents. *Pediatr Exerc Sci.* 1997;9:342–52.
342. Caspersen C, Nixon P, Durant R. Physical activity epidemiology applied to children and adolescents. *Exerc Sport Sci Rev.* 1998;26:341–403.
343. Oliver M, Schofield GM, Kolt GS. Physical activity in preschoolers: Understanding prevalence and measurement issues. *Sport Med.* 2007;37(12):1045–70.
344. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. *Med Care.* 2001;39(8):800–12.
345. Varni JW, Seid M, Knight TS, Uzark K, Szer IS. The PedsQL™ 4.0 Generic Core Scales: Sensitivity, Responsiveness, and Impact on Clinical Decision-Making. *J Behav Med.* 2002;25(2):175–93.
346. Bauman A, Ainsworth BE, Bull F, Craig CL, Hagströmer M, Sallis JF, et al. Progress and pitfalls in the use of the international physical activity questionnaire (IPAQ) for adult physical activity surveillance. *J Phys Act Heal.* 2009;6(Suppl 1):S5-8.

347. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res.* 2009;18(9):1263–78.
348. Bringolf-Isler B, Grize L, Mäder U, Ruch N, Sennhauser FH, Braun-Fahrländer C. Built environment, parents' perception, and children's vigorous outdoor play. *Prev Med (Baltim).* 2010;50(5–6):251–6.
349. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Vol. 37, *British Journal of Sports Medicine.* 2003. p. 197–206.
350. Linstone H, Turoff M. *The Delphi method: techniques and applications.* Reading, Mass.: Addison-Wesley Pub. Co.; 1975.
351. Powell C. The Delphi technique: Myths and realities. *J Adv Nurs.* 2003;41(4):376–82.
352. Almenara JC, Cejudo M del CL. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *EduTec Rev Electrónica Tecnol Educ.* 2013;7(2):11–22.
353. Ruiz-Olebuénaga J, Ispizua-Uribarri M. *La descodificación de la vida cotidiana.* Deustuko Unibertsitatea AZ, editor. 1998.
354. García-Abreu L, Fernández-García SJ. Procedimiento de aplicación del trabajo creativo en grupo de expertos. *Ing Energética.* 2011;29(2):46–50.
355. Landeta J. *El método Delphi: una técnica de previsión para la incertidumbre.* Barcelona: Editorial Ariel; 1999.
356. Martínez-Piñeiro E. La técnica Delphi como estrategia de consulta a los implicados en la evaluación de programas. *Rev Investig Educ.*

2003;21(2):449-63.

357. Vogt DS, King DW, King LA. Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychol Assess.* 2004;16(3):231-43.
358. Willis G. Cognitive interviewing. A tool for improving questionnaire design. Thousand O. 2005.
359. Blair J, Conrad F. Sample size for cognitive interview pretesting. *Public Opin Q.* 2011;75:636-658.
360. Kerr C, Nixon A, Wild D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Rev Pharmacoeconomics Outcomes Res.* 2010;10(3):269-81.
361. Beatty PC, Willis GB. Research synthesis: The practice of cognitive interviewing. *Public Opin Q.* 2007;71(2):287-311.
362. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrume. *Value Heal.* 2011;14(8):967-77.
363. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS. Patient-reported outcomes: Instrument development and selection issues. *Value Heal.* 2007;10(Suppl 2):S86-93.
364. Cattell R. The scientific use of factor analysis in behavioral and life sciences. New York, Plenum Press; 1978.
365. Everitt BS. Multivariate analysis: the need for data, and other problems. *Br J Psychiatry.* 1975;126(3):237-40.

366. Gorsuch RL. Exploratory Factor Analysis. Handbook of Multivariate Experimental Psychology. Boston, MA: Springer US; 1988.
367. Oliver M, Badland HM, Schofield GM, Shepherd J. Identification of accelerometer nonwear time and sedentary behavior. *Res Q Exerc Sport*. 2011;82(4):779–83.
368. Winkler EA, Gardiner PA, Healy GN, Clark BK, Sugiyama T, Matthews CE, et al. Distinguishing True Sedentary From Accelerometer Non-wearing Time: Accuracy Of Two Automated Wear-time Estimations. *Med Sci Sport Exerc*. 2009;41:171–2.
369. Troiano RP. Large-scale applications of accelerometers: New frontiers and new questions. *Med Sci Sports Exerc*. 2007;39(9):1501.
370. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoun L. Using accelerometers in youth physical activity studies: a review of methods. *J Phys Act Health*. 2013;10(3):437–50.
371. Janssen X, Basterfield L, Parkinson KN, Pearce MS, Reilly JK, Adamson AJ, et al. Objective measurement of sedentary behavior: Impact of non-wear time rules on changes in sedentary time. *BMC Public Health*. 2015;15(1):504.
372. Chinapaw MJM, De Niet M, Verloigne M, De Bourdeaudhuij I, Brug J, Altenburg TM. From sedentary time to sedentary patterns: Accelerometer data reduction decisions in youth. *PLoS One*. 2014;9(11):e111205.
373. Aadland E, Andersen LB, Anderssen SA, Resaland GK. A comparison of 10 accelerometer non-wear time criteria and logbooks in children. *BMC Public Health*. 2018;18(1):693–698.
374. International Children’s Accelerometry Database (ICAD). MRC

Epidemiology Unit, University of Cambridge. [Internet]. Available from:
<http://www.mrc-epid.cam.ac.uk/Research/Studies/ICAD/>

375. Toftager M, Kristensen PL, Oliver M, Duncan S, Christiansen LB, Boyle E, et al. Accelerometer data reduction in adolescents: Effects on sample retention and bias. *Int J Behav Nutr Phys Act.* 2013;10:140.
376. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sport Med.* 2017;47(9):1821–45.
377. Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, Almeida MJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc.* 2004;36(7):1259–66.
378. Mattocks C, Leary S, Ness A, Deere K, Saunders J, Tilling K, et al. Calibration of an accelerometer during free-living activities in children. *Int J Pediatr Obes.* 2007;2(4):218–26.
379. Pulsford RM, Cortina-Borja M, Rich C, Kinnafick FE, Dezateux C, Griffiths LJ. Actigraph Accelerometer-Defined boundaries for sedentary behaviour and physical activity intensities in 7 year old children. *PLoS One.* 2011;6(8):e21822.
380. Johansson E, Ekelund U, Nero H, Marcus C, Hagströmer M. Calibration and cross-validation of a wrist-worn Actigraph in young preschoolers. *Pediatr Obes.* 2015;10(1):1–6.
381. Butte NF, Wong WW, Lee JS, Adolph AL, Puyau MR, Zakeri IF. Prediction of energy expenditure and physical activity in preschoolers. *Med Sci Sports Exerc.* 2014;46(6):1216–26.
382. Trost SG, Fees BS, Haar SJ, Murray AD, Crowe LK. Identification and

- validity of accelerometer cut-points for toddlers. *Obesity*. 2012;20(11):2317–9.
383. Williams R. Kcal Estimates from Activity Counts using the Potential Energy Method. CSA, Inc, Actigraph Corp. 1998;
384. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777–81.
385. What is the difference among the Energy Expenditure Algorithms? [Internet]. Available from: <https://actigraphcorp.force.com/support/s/article/What-is-the-difference-among-the-Energy-Expenditure-Algorithms>
386. Tudor-Locke C, Johnson WD, Katzmarzyk PT. U.S. population profile of time-stamped accelerometer outputs: Impact of wear time. *J Phys Act Heal*. 2011;8(5):693–8.
387. Benítez-Porres J, Alvero-Cruz JR, Sardinha LB, López-Fernández I, Carnero EA. Cut-off values for classifying active children and adolescents using the Physical Activity Questionnaire: PAQ-C and PAQ-A. *Nutr Hosp*. 2016;33(5):1036–44.
388. Chen S-R, Lee Y-J, Chiu H-W, Jeng C. Impact of physical activity on heart rate variability in children with type 1 diabetes. *Child's Nerv Syst ChNS Off J Int Soc Pediatr Neurosurg*. 2008 Jun;24(6):741–7.
389. Ogunleye AA, Voss C, Sandercock GR. Prevalence of high screen time in English youth: Association with deprivation and physical activity. *J Public Health (Bangkok)*. 2012;34(1):46–53.
390. Bailey DA, McKay HA, Mirwald RL, Crocker PRE, Faulkner RA. A six-year longitudinal study of the relationship of physical activity to bone

mineral accrual in growing children: The University of Saskatchewan
Bone Mineral Accrual Study. *J Bone Miner Res.* 1999;14(10):1672-9.

8. Anexos

8.1. Anexo 1. Riesgo de sesgo de COSMIN. Versión 2018

Fiabilidad

A. Requerimientos de diseño

	Excelente	Adecuado	Dudoso	Inadecuado	NA
1. ¿Los sujetos se mantuvieron estables en el período intermedio en el constructo a medir?	Se proporcionó evidencia de que los sujetos se mantuvieron estables	Se asume que los sujetos se mantuvieron estables	No está claro si los sujetos se mantuvieron estables	Los sujetos no se mantuvieron estables	
2. ¿Fue apropiado el intervalo de tiempo?	Intervalo de tiempo apropiado		Dudoso si el intervalo de tiempo fue apropiado o no estaba fijado	El intervalo de tiempo no fue apropiado	
3. ¿Fueron las condiciones de prueba similares para las mediciones? p.ej. tipo de administración, entorno, instrucciones	Se proporcionó evidencia de que las condiciones del test fueron similares	Se asume que las condiciones del test fueron similares	No está claro si las condiciones del test fueron similares	Las condiciones del test no fueron similares	

B. Métodos estadísticos

	Excelente	Adecuado	Dudoso	Inadecuado	NA
4. Para puntuaciones continuas: ¿Se calculó el índice de correlación intraclase (ICC)?	ICC calculado y se describe el modelo o fórmula de ICC	ICC calculado, pero no se describe el modelo o la fórmula de ICC o no es óptimo. Coeficiente de correlación de Pearson o Spearman calculado con evidencia siempre que no haya ocurrido ningún cambio sistemático	Coeficiente de correlación de Pearson o Spearman calculado SIN evidencia siempre que no se haya producido un cambio sistemático o CON evidencia de que se ha producido un cambio sistemático	No se calcularon correlaciones de ICC, Pearson o Spearman	No aplicable
5. Para puntuaciones dicotómicas / nominales / ordinales: ¿Se calculó kappa?	Kappa calculado			Kappa no calculado	No aplicable
6. Para puntuaciones ordinales: ¿Se calculó un kappa ponderado?	Kappa ponderado calculado		Kappa ponderado no calculado o no descrito		No aplicable
7. Para puntuaciones ordinales: ¿Se describió el esquema de ponderación? p.ej. lineal, cuadrático	Esquema de ponderación descrito	Esquema de ponderación no descrito			No aplicable

C. Otros

	Excelente	Adecuado	Dudoso	Inadecuado	NA
8. ¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?	No hubo otros fallos metodológicos importantes		Otros fallos metodológicos menores	Otros fallos metodológicos importantes	

Error de medición

A. Requerimientos de diseño

	Excelente	Adecuado	Dudoso	Inadecuado	NA
1. ¿Los sujetos se mantuvieron estables en el período intermedio en el constructo a medir?	Se proporcionó evidencia de que los sujetos se mantuvieron estables	Se asume que los sujetos se mantuvieron estables	No está claro si los sujetos se mantuvieron estables	Los sujetos no se mantuvieron estables	
2. ¿Fue apropiado el intervalo de tiempo?	Intervalo de tiempo apropiado		Dudoso si el intervalo de tiempo fue apropiado o no estaba fijado	El intervalo de tiempo no fue apropiado	
3. ¿Fueron las condiciones de prueba similares para las mediciones? p.ej. tipo de administración, entorno, etc.	Se proporcionó evidencia de que las condiciones del test fueron similares	Se asume que las condiciones del test fueron similares	No está claro si las condiciones del test fueron similares	Las condiciones del test no fueron similares	

B. Métodos estadísticos

	Excelente	Adecuado	Dudoso	Inadecuado	NA
4. Para puntuaciones continuas: ¿Se calculó el error estándar de medición (SEM), el cambio más pequeño detectable (SDC) o los límites de acuerdo (LoA)?	SEM, SDC o LoA calculado	Es posible calcular el LoA a partir de los datos presentados		SEM calculado en base al alfa de Cronbach o en SD de otra población	No aplicable
5. Para puntuaciones dicotómicas / nominales / ordinales: ¿Se calculó el porcentaje de acuerdo (PoA) (positiva y negativa)?	PoA positivo y negativo calculado	PoA calculado		PoA no calculado	No aplicable

C. Otros

	Excelente	Adecuado	Dudoso	Inadecuado	NA
6. ¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?	No hubo otros fallos metodológicos importantes		Otros fallos metodológicos menores	Otros fallos metodológicos importantes	

Test de hipótesis para la validez de constructo

Comparación con otros instrumentos de medición de resultados (validez convergente)

A. Requerimientos de diseño

	Excelente	Adecuado	Dudoso	Inadecuado	NA
1. ¿Está claro qué miden el (los) instrumento (s) de comparación?	Los constructos por el (los) instrumento (s) comparador son claros			Los constructos medidos por el (los) instrumento (s) comparador no están claros	
2. ¿Fueron suficientes las propiedades de medición del (de los) instrumento (s) comparador?	Propiedades de medición suficientes del (de los) instrumento (s) comparador en una población similar a la del estudio	Propiedades de medición suficientes del (de los) instrumento (s) de comparación, pero no estoy seguro de si se aplican a la población de estudio	Alguna información sobre las propiedades de medición de los instrumentos comparadores en cualquier población de estudio	No hay información sobre las propiedades de medición del (de los) instrumento (s) comparador, O evidencia de propiedades de medición insuficientes del (de los) instrumento (s) comparador	

B. Métodos estadísticos

	Excelente	Adecuado	Dudoso	Inadecuado	NA
3. ¿Fue el método estadístico apropiado para probar las hipótesis?	El método estadístico fue apropiado	Se supone que el método estadístico es apropiado	Método estadístico aplicado NO óptimo	Método estadístico aplicado NO apropiado	

C. Otros

	Excelente	Adecuado	Dudoso	Inadecuado	NA
4. ¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?	No hubo otros fallos metodológicos importantes		Otros fallos metodológicos menores	Otros fallos metodológicos importantes	

Comparación entre subgrupos (discriminativo o conocido – validez de grupos conocidos)

A. Requerimientos de diseño

	Excelente	Adecuado	Dudoso	Inadecuado	NA
5. ¿Se proporcionó una descripción adecuada de las características importantes de los subgrupos?	Descripción adecuada de las características importantes de los subgrupos	Descripción adecuada de la mayoría de las características importantes de los subgrupos.	Deficiente o sin descripción de las características importantes de los subgrupos		

B. Métodos estadísticos

	Excelente	Adecuado	Dudoso	Inadecuado	NA
6. ¿Fue el método estadístico apropiado para probar las hipótesis?	El método estadístico fue apropiado	Se supone que el método estadístico es apropiado	Método estadístico aplicado NO óptimo	Método estadístico aplicado NO apropiado	

C. Otros

	Excelente	Adecuado	Dudoso	Inadecuado	NA
7. ¿Hubo otros fallos importantes en el diseño o los métodos estadísticos del estudio?	No hubo otros fallos metodológicos importantes		Otros fallos metodológicos menores	Otros fallos metodológicos importantes	

8.2. Anexo 2. Versión original del EY-PAQ

Study ID: _____
APPENDIX A



BRADFORD INSTITUTE
FOR HEALTH RESEARCH



Early Years Physical Activity Questionnaire

Please complete this questionnaire when your child has worn their physical activity monitor for 7 days. The questionnaire asks about the types of activities your child has been doing in the last 4 weeks, how often and for how long. There are three parts. There are no right or wrong answers and all responses are confidential. Please answer all questions, unless there are questions that you do not wish to answer, you can skip these. If you are unsure about any of the questions please ask a member of school staff or the research team for guidance when you return the questionnaire and activity monitors.

Example answers... In the last month, how many days each week and for how long each day would you say your child has spent doing the following activities at home?

	Less than once a week	How often	For how long each day Please tick one box or write a time if more than one hour daily			
	Yes/ No	Number of days each week. If never put a zero	Up to 15 mins/day	16-30 mins/day	31-60 mins/day	More than an hour/day. Please estimate time
Colouring/drawing/crafting	No	5/7				Hrs: 2 Mins: 30
Sitting playing with toys (e.g. dolls/puzzles/educational play)	No	7/7			✓	Hrs:..... Mins:.....
Watching TV/DVDs	No	7/7				Hrs: 1 Mins: 45
Playing on the computer (do not include physically active games such as Nintendo Wii)	Yes/7		✓		Hrs:..... Mins:.....

Thank you in advance for completing the questionnaire

Questionnaire completed by (please circle): Mother / Father / Other (please specify):

Q1. In the last month, how many days each week and for how long each day would you say your child has spent doing the following activities at home?

	Less than once a week	How often	For how long each day Please tick one box or write a time if more than one hour daily			
	Yes/ No	Number of days each week. If never put a zero	Up to 15 mins/day	16-30 mins/day	31-60 mins/day	More than an hour/day, please estimate time
Colouring/drawing/Craft	/7				Hrs:..... Mins:.....
Sitting playing with toys (e.g. dolls/puzzles/educational play)	/7				Hrs:..... Mins:.....
Watching TV/DVDs	/7				Hrs:..... Mins:.....
Playing on the computer (do not include physically active games such as Nintendo Wii)	/7				Hrs:..... Mins:.....
Sitting listening/singing to music	/7				Hrs:..... Mins:.....
Reading/being read to	/7				Hrs:..... Mins:.....
Playing actively inside the house (e.g. dancing, crawling, running, sit and ride toys, push toys, physically active computer games such as Nintendo Wii)	/7				Hrs:..... Mins:.....
Playing actively in the garden/yard	/7				Hrs:..... Mins:.....
Engaging in physical activity/active play that makes them sweat or breathe harder	/7				Hrs:..... Mins:.....

Q2. In the last month, to get from place to place (e.g to the shops, school/groups, park, visiting friends/relatives), on how many days each week and for how long each day would you say your child has spent:

	Less than once a week	How often	For how long each day Please tick one box or write a time if more than one hour daily			
	Yes/ No	Number of days each week. If never put a zero	Up to 15 mins/day	16-30 mins/day	31-60 mins/day	More than an hour/day, please estimate time
In their buggy/pushchair	/7				Hrs:..... Mins:.....
Walking	/7				Hrs:..... Mins:.....
Being carried	/7				Hrs:..... Mins:.....
In the car	/7				Hrs:..... Mins:.....
On public transport	/7				Hrs:..... Mins:.....

Q3. In the last month, on how many days each week and for how long each day has your child done the following?

	Less than once a week	How often	For how long each day Please tick one box or write a time if more than one hour daily			
	Yes/ No	Number of days each week. If never put a zero	Up to 15 mins/day	16-30 mins/day	31-60 mins/day	More than an hour/day, please estimate time
Played at the park/playground	/7				Hrs:..... Mins:.....
Played at Indoor play facilities (e.g. ball pools)	/7				Hrs:..... Mins:.....

8.3. Anexo 3. Versión original del PAQ-C

Physical Activity Questionnaire (Elementary School)

Name: _____

Age: _____

Sex: M _____ F _____

Grade: _____

Teacher: _____

We are trying to find out about your level of physical activity from *the last 7 days* (in the last week). This includes sports or dance that make you sweat or make your legs feel tired, or games that make you breathe hard, like tag, skipping, running, climbing, and others.

Remember:

1. There are no right and wrong answers — this is not a test.
2. Please answer all the questions as honestly and accurately as you can — this is very important.

1. Physical activity in your spare time: Have you done any of the following activities in the past 7 days (last week)? If yes, how many times? (Mark only one circle per row.)

	No	1-2	3-4	5-6	7 times or more
Skipping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rowing/canoeing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In-line skating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tag	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Walking for exercise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bicycling	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jogging or running	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aerobics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Swimming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Baseball, softball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Football	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Badminton	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skateboarding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Soccer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Street hockey	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Volleyball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Floor hockey	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basketball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ice skating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cross-country skiing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ice hockey/ringette	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other:					
_____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
_____	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. In the last 7 days, during your physical education (PE) classes, how often were you very active (playing hard, running, jumping, throwing)? (Check one only.)

- I don't do PE
- Hardly ever
- Sometimes
- Quite often
- Always

3. In the last 7 days, what did you do most of the time *at recess*? (Check one only.)

- Sat down (talking, reading, doing schoolwork).....
- Stood around or walked around
- Ran or played a little bit
- Ran around and played quite a bit
- Ran and played hard most of the time

4. In the last 7 days, what did you normally do *at lunch* (besides eating lunch)? (Check one only.)

- Sat down (talking, reading, doing schoolwork).....
- Stood around or walked around
- Ran or played a little bit
- Ran around and played quite a bit
- Ran and played hard most of the time

5. In the last 7 days, on how many days *right after school*, did you do sports, dance, or play games in which you were very active? (Check one only.)

- None
- 1 time last week
- 2 or 3 times last week
- 4 times last week
- 5 times last week

6. In the last 7 days, on how many *evenings* did you do sports, dance, or play games in which you were very active? (Check one only.)

- None
- 1 time last week
- 2 or 3 times last week
- 4 or 5 last week
- 6 or 7 times last week

7. On the last weekend, how many times did you do sports, dance, or play games in which you were very active? (Check one only.)

- None
- 1 time
- 2 — 3 times
- 4 — 5 times
- 6 or more times

8. Which *one* of the following describes you best for the last 7 days? Read *all five* statements before deciding on the *one* answer that describes you.

- A. All or most of my free time was spent doing things that involve little physical effort
- B. I sometimes (1 — 2 times last week) did physical things in my free time (e.g. played sports, went running, swimming, bike riding, did aerobics)
- C. I often (3 — 4 times last week) did physical things in my free time
- D. I quite often (5 — 6 times last week) did physical things in my free time
- E. I very often (7 or more times last week) did physical things in my free time

9. Mark how often you did physical activity (like playing sports, games, doing dance, or any other physical activity) for each day last week.

	None	Little bit	Medium	Often	Very often
Monday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tuesday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wednesday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Thursday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Saturday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sunday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Were you sick last week, or did anything prevent you from doing your normal physical activities? (Check one.)

- Yes
- No

If Yes, what prevented you? _____

8.4. Anexo 4. Aprobación del comité de ética de investigación

APROVACIÓ PROJECTE PEL CER/ APROBACIÓN PROYECTO POR EL CER

Codi de l'estudi / Código del estudio: FIS-2018-06

Versió del protocol / Versión del protocolo: 1.0

Data de la versió / Fecha de la versión: 20/11/18

Sant Cugat del Vallès, 8 de juliol de 2019

Equip/o investigador: Marta Amor Barbosa (IP), Dra. Montserrat Girabent Farrés, Dra. Caritat Bagur Calafat

Títol de l'estudi / Título del estudio: Validez y fiabilidad del cuestionario Physical Activity Questionnaire – Young Children (PAQ-YC) para niños entre 5-7 años.

Benvolgut/da,

Valorat el projecte presentat, el CER de la Universitat Internacional de Catalunya, considera que, el contingut de la investigació, no implica cap inconvenient relacionat amb la dignitat humana, tracte ètic per als animals ni atempta contra el medi ambient, ni té implicacions econòmiques ni conflicte d'interessos, **però no s'han valorat els aspectes metodològics del projecte de recerca degut a que tal anàlisis correspon a d'altres instàncies.**

Per aquests motius, el Comitè d'Ètica de Recerca, **RESOLT FAVORABLEMENT**, emetre aquest **CERTIFICAT D'APROVACIÓ**, per que pugui ser presentat a les instàncies que així ho requereixin.

Em permeto recordar-li que, si en el procés d'execució es produís algun canvi significatiu en els seus plantejaments, hauria de ser sotmès novament a la revisió i aprovació del CER.

Atentament,

Apreciado/a,

Valorado el proyecto presentado, el CER de la Universidad Internacional de Catalunya, considera que, el contenido de la investigación, no implica ningún inconveniente relacionado con la dignidad humana, trato ético para los animales, ni atenta contra el medio ambiente, ni tiene implicaciones económicas ni conflicto de intereses, pero no se han valorado aspectos metodológicos del proyecto de investigación debido a que tal análisis corresponde a otras instancias.

Por estos motivos, el Comité d'Ètica de Recerca, RESUELVE FAVORABLEMENTE, emitir este CERTIFICADO DE APROBACIÓN, para que pueda ser presentado a las instancias que así lo requieran.

Me permito recordarle que, si el proceso de ejecución se produjera algún cambio significativo en sus planteamientos, debería ser sometido nuevamente a la revisión y aprobación del CER.

Atentamente,



Dr. Josep Maria Guardiola
President CER-UIC

8.5. Anexo 5. Documento de información y consentimiento informado del estudio de fiabilidad



DOCUMENTO DE INFORMACIÓN AL SUJETO PARTICIPANTE DEL ESTUDIO DE INVESTIGACIÓN

Código del protocolo de investigación:
Versión del protocolo: 1
Fecha de la versión del protocolo: 21/11/2018
Fecha de la presentación del protocolo:
Título del Proyecto: Validez y fiabilidad del cuestionario Physical Activity Questionnaire – Young Children (PAQ-YC) para niños entre 5-7 años.

Director/a del Proyecto: Dra. Montserrat Girabent Farrés Investigador/a: Marta Amor Barbosa Departamento: Fisioterapia
--

Hemos solicitado su participación en un estudio de investigación. Antes de decidir si aceptan participar, es importante que comprendan los motivos por los cuales se lleva a cabo la investigación: como se usará su información, en qué consistirá el estudio y los posibles beneficios, riesgos y molestias que pueda comportar.

En caso de que participen en algún otro estudio, lo tendrán que comunicar al responsable para valorar si pueden participar en este.

¿CUÁLES SON LOS ANTECEDENTES Y EL OBJETIVO DE ESTE ESTUDIO?

La actividad física se relaciona con un mejor estado de salud, bienestar psicológico y social, prevención de enfermedades e incluso un mejor rendimiento académico. Por ello, es de gran valor potenciar la actividad física desde edades tempranas para implementarla como un hábito de vida saludable.

Existen diferentes instrumentos para medir la actividad física en la infancia. Se han diseñado dispositivos electrónicos bastante complejos, costosos, poco prácticos y, en general, de difícil acceso. Como alternativa a estos dispositivos, han surgido los cuestionarios, que son instrumentos asequibles, de fácil aplicación y que además son válidos para medir la actividad física en niños.

Nos ponemos en contacto con usted para proponerle participar en un proyecto de validación de un cuestionario de actividad física que se ha creado específicamente para niños/as entre 5 y 7 años. Su colaboración consistirá en rellenar el cuestionario en dos ocasiones.

El cuestionario está diseñado de forma que las preguntas se refieren a la actividad física que su hijo/a ha hecho durante los últimos 7 días. Por tanto, una de las cosas se han de valorar es la concordancia en sus respuestas. Esto quiere decir que, si les preguntan por los mismos 7 días en diferentes momentos, sus respuestas no tendrían que cambiar.

Por ese motivo, les pedimos que cubran el cuestionario dos veces, durante dos días seguidos, pero respondiendo a las preguntas sobre los mismos 7 días en las dos ocasiones. Si por cualquier motivo, no pudieran responder al cuestionario al día siguiente, pueden hacerlo al cabo de dos o incluso tres días, pero siempre teniendo en cuenta que deben recordar los mismos 7 días.

¿TENGO LA OBLIGACIÓN DE PARTICIPAR?

La decisión sobre participar o no en la investigación corresponde a ustedes. Si deciden participar, les pasaremos un formulario de consentimiento informado para que lo firmen.

¿CUÁLES SON MIS OBLIGACIONES?

Su obligación es cubrir el cuestionario en las dos ocasiones que se solicita, cumpliendo los plazos de tiempo que se han descrito anteriormente y respondiendo con total sinceridad a las preguntas que se plantean.

¿CUÁLES SON LOS POSIBLES EFECTOS SECUNDARIOS, RIESGOS Y MOLESTIAS ASOCIADOS A LA PARTICIPACIÓN?

No existe ningún efecto secundario, riesgo o molestia asociado a la participación.

¿CUÁLES SON LOS POSIBLES BENEFICIOS DE PARTICIPAR?

La actividad física es uno de los pilares básicos en la prevención de enfermedades crónicas como la obesidad, patologías cardiovasculares y síndromes metabólicos. Por tanto, es importante disponer de instrumentos válidos que permitan medirla y que cuantifiquen su impacto en la salud desde edades tempranas.

Su participación en el estudio permitirá obtener la versión de un cuestionario que valora la actividad física en niños entre 5-7 años.

¿CÓMO SE UTILIZARÁN MIS DATOS EN EL ESTUDIO?

El trato, la comunicación y la cesión de los datos de carácter personal de los sujetos participantes en el ensayo se ajustan a lo que dispone la Ley orgánica 3/2018 del 5 de diciembre y normativa aplicable en esta materia, de protección de datos de carácter personal.

Estos datos, no incluyen ni su nombre ni su dirección, sino que se le asignará un número de código. Únicamente el equipo investigador, tendrá acceso a la clave del código que permite asociar los datos del estudio con ustedes. No obstante, las autoridades reguladoras, el comité de ética independiente u otras entidades de supervisión podrán revisar sus datos personales. El objetivo de dichas revisiones es garantizar la dirección adecuada del estudio o la calidad de los datos del estudio.

Si retiran del consentimiento informado de usar sus datos para el estudio, no podrán continuar participando en la investigación. Han de tener en cuenta que los resultados del estudio pueden aparecer publicados en la bibliografía, si bien, su identidad no será revelada.

¿CÓMO PUEDO ESTABLECER CONTACTO SI NECESITO OBTENER MÁS INFORMACIÓN O AYUDA?

Mediante la firma de este formulario, asienten que han sido informados de las características del estudio, han entendido la información y se les ha clarificado todas sus dudas.

En caso de padecer un daño relacionado con el estudio o para obtener respuesta a cualquier pregunta que pueda surgir durante la investigación contacte con:

Sra. Marta Amor Barbosa
Universitat Internacional de Catalunya
Dirección: C/ Josep Trueta, s/n, 08195, Sant Cugat del Vallès
Nº de teléfono: 93 504 20 00

CONSENTIMIENTO INFORMADO DOCUMENTO DE INFORMACIÓN AL SUJETO

PARTICIPANTE DEL ESTUDIO DE INVESTIGACIÓN

CONSENTIMIENTO INFORMADO

Código del estudio:

Versión del protocolo: 1

Fecha de la versión: 20/11/2018

Fecha de la presentación:

Título del Proyecto: Validez y fiabilidad del cuestionario Physical Activity Questionnaire – Young Children (PAQ-YC) para niños entre 5-7 años.

Director/a del Proyecto: Dra. Montserrat Girabent Farrés

Investigador/a: Marta Amor Barbosa

Departamento: Fisioterapia

Yo, el Sr./la Sra:

- He recibido información verbal sobre el estudio y he leído la información escrita que se adjunta, la cual me ha sido facilitada una copia.
- He comprendido lo que se me ha explicado y los posibles riesgos y beneficios de participar en el estudio.
- He podido comentar el estudio y hacer preguntas al profesional responsable.
- Doy mi consentimiento para tomar parte en el estudio y asumo que mi participación es totalmente voluntaria.
- Entiendo que me podré retirar en cualquier momento.

Mediante la firma de este formulario de consentimiento informado, doy mi para que mis datos personales se puedan usar como se ha descrito en este formulario de consentimiento, que se ajusta a lo que dispone la Ley orgánica 3/2018, de 5 de diciembre, de protección de datos de carácter personal.

Entiendo que recibiré una copia de este formulario de consentimiento informado.

Firma del Participante
Núm. de DNI

Fecha de la firma

Firma del Investigador/a
Nombre: Marta Amor Barbosa

Fecha de la firma



8.6. Anexo 6. Documento de información y consentimiento informado del estudio de validez de constructo



DOCUMENTO DE INFORMACIÓN AL SUJETO PARTICIPANTE DEL ESTUDIO DE INVESTIGACIÓN

Código del protocolo de investigación:

Versión del protocolo: 1

Fecha de la versión del protocolo: 21/11/2018

Fecha de la presentación del protocolo:

Título del Proyecto: Validez y fiabilidad del cuestionario Physical Activity Questionnaire – Young Children (PAQ-YC) para niños entre 5-7 años.

Director/a del Proyecto: Dra. Montserrat Girabent Farrés

Investigador/a: Marta Amor Barbosa

Departamento: Fisioterapia

Hemos solicitado su participación en un estudio de investigación. Antes de decidir si aceptan participar, es importante que comprendan los motivos por los cuales se lleva a cabo la investigación: como se usará su información, en qué consistirá el estudio y los posibles beneficios, riesgos y molestias que pueda comportar.

En caso de que participen en algún otro estudio, lo tendrán que comunicar al responsable para valorar si pueden participar en este.

¿CUÁLES SON LOS ANTECEDENTES Y EL OBJETIVO DE ESTE ESTUDIO?

La actividad física se relaciona con un mejor estado de salud, bienestar psicológico y social, prevención de enfermedades e incluso un mejor rendimiento académico. Por ello, es de gran valor potenciar la actividad física desde edades tempranas para implementarla como un hábito de vida saludable.

Existen diferentes instrumentos para medir la actividad física en la infancia. Se han diseñado dispositivos electrónicos bastante complejos, costosos, poco prácticos y, en general, de difícil acceso. Como alternativa a estos dispositivos, han surgido los cuestionarios, que son instrumentos asequibles, de fácil aplicación y que además son válidos para medir la actividad física en niños.

Nos ponemos en contacto con usted para proponerle participar en un proyecto de validación de un cuestionario de actividad física que se ha creado específicamente para niños/as entre 5 y 7 años. Su colaboración consistirá en supervisar que su hijo/a lleve un acelerómetro durante 7 días y rellenar el cuestionario de actividad física.

El cuestionario está diseñado de forma que las preguntas se refieren a la actividad física que su hijo/a ha hecho durante los últimos 7 días. Por ello, en primer lugar, se ha de registrar la actividad física que hace durante una semana. Esto se hará con un acelerómetro, que es un dispositivo que se coloca en la cintura y mide el movimiento y las aceleraciones. Pesa 20 gramos, es muy cómodo de llevar y no genera ninguna interferencia en la vida normal del niño/a. Una persona del equipo será la encargada de entregarlo y de recogerlo. En segundo lugar, cuando se retire el acelerómetro, deberá responder al cuestionario. Como las preguntas hacen referencia a la actividad que su hijo o hija ha hecho en los últimos 7 días, coincide con los días en que ha llevado puesto el acelerómetro. De esa forma, se podrá comprobar si hay

concordancia entre lo que registra el acelerómetro y las respuestas que usted da al cuestionario.

¿TENGO LA OBLIGACIÓN DE PARTICIPAR?

La decisión sobre participar o no en la investigación corresponde a ustedes. Si deciden participar, les pasaremos un formulario de consentimiento informado para que lo firmen.

¿CUÁLES SON MIS OBLIGACIONES?

Su obligación consiste en supervisar que su hijo/a lleve el acelerómetro durante 7 días, manejar el dispositivo con responsabilidad y responder el cuestionario una vez finalizado este período, cumpliendo los plazos de tiempo que se han descrito anteriormente y respondiendo con total sinceridad a las preguntas que se plantean.

¿CUÁLES SON LOS POSIBLES EFECTOS SECUNDARIOS, RIESGOS Y MOLESTIAS ASOCIADOS A LA PARTICIPACIÓN?

No existe ningún efecto secundario, riesgo o molestia asociado a la participación.

¿CUÁLES SON LOS POSIBLES BENEFICIOS DE PARTICIPAR?

La actividad física es uno de los pilares básicos en la prevención de enfermedades crónicas como la obesidad, patologías cardiovasculares y síndromes metabólicos. Por tanto, es importante disponer de instrumentos válidos que permitan medirla y que cuantifiquen su impacto en la salud desde edades tempranas.

Su participación en el estudio permitirá obtener la versión de un cuestionario que valora la actividad física en niños entre 5-7 años.

¿CÓMO SE UTILIZARÁN MIS DATOS EN EL ESTUDIO?

El trato, la comunicación y la cesión de los datos de carácter personal de los sujetos participantes en el ensayo se ajustan a lo que dispone la Ley orgánica 3/2018 del 5 de diciembre y normativa aplicable en esta materia, de protección de datos de carácter personal.

Estos datos, no incluyen ni su nombre ni su dirección, sino que se le asignará un número de código. Únicamente el equipo investigador, tendrá acceso a la clave del código que permite asociar los datos del estudio con ustedes. No obstante, las autoridades reguladoras, el comité de ética independiente u otras entidades de supervisión podrán revisar sus datos personales. El objetivo de dichas revisiones es garantizar la dirección adecuada del estudio o la calidad de los datos del estudio.

Si retiran del consentimiento informado de usar sus datos para el estudio, no podrán continuar participando en la investigación. Han de tener en cuenta que los resultados del estudio pueden aparecer publicados en la bibliografía, si bien, su identidad no será revelada.

¿CÓMO PUEDO ESTABLECER CONTACTO SI NECESITO OBTENER MÁS INFORMACIÓN O AYUDA?

Mediante la firma de este formulario, asienten que han sido informados de las características del estudio, han entendido la información y se les ha clarificado todas sus dudas.

En caso de padecer un daño relacionado con el estudio o para obtener respuesta a cualquier pregunta que pueda surgir durante la investigación contacte con:

Sra. Marta Amor Barbosa
Universitat Internacional de Catalunya
Dirección: C/ Josep Trueta, s/n, 08195, Sant Cugat del Vallès
Nº de teléfono: 93 504 20 00

CONSENTIMIENTO INFORMADO DOCUMENTO DE INFORMACIÓN AL SUJETO

PARTICIPANTE DEL ESTUDIO DE INVESTIGACIÓN

CONSENTIMIENTO INFORMADO

Código del estudio:

Versión del protocolo: 1

Fecha de la versión: 20/11/2018

Fecha de la presentación:

Título del Proyecto: Validez y fiabilidad del cuestionario Physical Activity Questionnaire – Young Children (PAQ-YC) para niños entre 5-7 años.

Director/a del Proyecto: Dra. Montserrat Girabent Farrés
 Investigador/a: Marta Amor Barbosa
 Departamento: Fisioterapia

Yo, el Sr./la Sra:

- He recibido información verbal sobre el estudio y he leído la información escrita que se adjunta, la cual me ha sido facilitada una copia.
- He comprendido lo que se me ha explicado y los posibles riesgos y beneficios de participar en el estudio.
- He podido comentar el estudio y hacer preguntas al profesional responsable.
- Doy mi consentimiento para tomar parte en el estudio y asumo que mi participación es totalmente voluntaria.
- Entiendo que me podré retirar en cualquier momento.

Mediante la firma de este formulario de consentimiento informado, doy mi para que mis datos personales se puedan usar como se ha descrito en este formulario de consentimiento, que se ajusta a lo que dispone la Ley orgánica 3/2018, de 5 de diciembre, de protección de datos de carácter personal.

Entiendo que recibiré una copia de este formulario de consentimiento informado.

 Firma del Participante
 Núm. de DNI

 Fecha de la firma

 Firma del Investigador/a
 Nombre: Marta Amor Barbosa

 Fecha de la firma



8.7. Anexo 7. Development and Content Validity of the Physical Activity Questionnaire-Young Children (PAQ-YC) to Assess Physical Activity in Children between 5 and 7 Years



Article

Development and Content Validity of the Physical Activity Questionnaire-Young Children (PAQ-YC) to Assess Physical Activity in Children between 5 and 7 Years

Marta Amor-Barbosa ¹, Montserrat Girabent-Farrés ^{2,*}, Ferran Rosés-Noguer ^{3,4}, Anna Ortega-Martínez ^{1,5}, Almudena Medina-Rincón ^{2,6} and Caritat Bagur-Calafat ¹

- ¹ Physiotherapy Department, Universitat Internacional de Catalunya, 08195 Barcelona, Spain; mamor@uic.es (M.A.-B.); aortegam@uic.es (A.O.-M.); cbagur@uic.es (C.B.-C.)
 - ² Physiotherapy Department, School of Health Sciences, TecnoCampus-Pompeu Fabra University, Mataró, 08302 Barcelona, Spain; almunrincon@hotmail.com
 - ³ Paediatric Cardiology Department, Vall d'Hebron University Hospital, 08035 Barcelona, Spain; froses@vhebron.net
 - ⁴ Paediatric Cardiology Department, Royal Brompton and Harefield NHS Foundation Trust, London SW3 6NR, UK
 - ⁵ Physiotherapy Department, Fundació Aspace Catalunya, 08001 Barcelona, Spain
 - ⁶ Rehabilitation Section, Centro Hospitalario Pere Virgili, 08023 Barcelona, Spain
- * Correspondence: mgirabent@tecnocampus.cat



Citation: Amor-Barbosa, M.; Girabent-Farrés, M.; Rosés-Noguer, F.; Ortega-Martínez, A.; Medina-Rincón, A.; Bagur-Calafat, C. Development and Content Validity of the Physical Activity Questionnaire-Young Children (PAQ-YC) to Assess Physical Activity in Children between 5 and 7 Years. *Healthcare* **2021**, *9*, 655. <https://doi.org/10.3390/healthcare9060655>

Academic Editor: Filipe Manuel Clemente

Received: 30 April 2021
Accepted: 28 May 2021
Published: 31 May 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: Childhood is a critical period in the development and consolidation of healthy habits, such as the practice of physical activity (PA). It is essential to have valid instruments to measure PA from an early age. The aim of this study was to design and evaluate the content validity of the Physical Activity Questionnaire-Young Children (PAQ-YC) to measure the PA level in children aged 5–7 years. The first version of the questionnaire was tested by a 2-round Delphi study. It was established as a consensus criterion that the relative interquartile range (RIR) and/or the coefficient of variation (CV) were $\leq 20\%$. The most significant discrepancies in the Delphi survey ($n = 11$ – 13) were observed for items about hours of Physical Education or similar activities at school (item 7: RIR = 20, CV = 38.73) and for items about participation in Physical Education (item 8: RIR = 25, CV = 15.45). The cognitive interviews ($n = 5$) confirmed the version agreed by the experts. The results show that the PAQ-YC presents adequate content validity in terms of relevance, comprehensiveness and comprehensibility.

Keywords: physical activity; questionnaire; content validity; paediatrics

1. Introduction

Physical activity (PA) is a fundamental basis in chronic disease prevention, and it is related to numerous benefits apart from physical health, such as psychological well-being or cognitive performance. Solid evidence demonstrates that PA is associated with a reduced risk of increasing weight and adiposity [1–3] and favourable signs in bone health, such as bone structural strength, bone mineral content and area [4–6] in preschoolers. For children and adolescents, it has been proved that regular PA is related to better cardiorespiratory resistance [7–11], better muscular resistance and strength [12–15], decreased weight and corporal fat percentage [16–18], better bone health [17,19–21], less risk of cardiovascular and metabolic diseases [11,22,23], reduction of depressive symptoms or anxiety [24–27] and an improvement on cognitive indicators, such as memory, processing velocity, attention and academic performance [28–30]. Considering the impact of sedentary conduct on the health of children and adolescents, evidence suggests that the higher percentage of sedentarism, the less cardiovascular and metabolic health and higher adiposity index [31–33].

Childhood is a critical period in developing and consolidating healthy habits since a child's brain is sensitive to being influenced by the environment in its modifiable conduct [34]. Therefore, it is essential to have valid instruments capable of objectifying the

level of PA performed from an early age. Even though some authors consider that doubly labelled water [35], calorimetry [36] or direct observation [37] are the gold standards in measuring levels of PA in children, others defend that there is no real gold standard [38–43]. Scientific evidence suggests that any current method can capture all the components of PA, and the selection decision can only rely on objective criteria depending on practicality, psychometric properties and the aim of the investigation [41,44–46].

Compared to direct methods, questionnaires offer some advantages, such as low cost and associated time of distribution and administration, and the possibility of gathering information from many participants [45,47]. Another strength is that they can identify dimensions and domains of PA [38,44] and have no risk of affecting usual patterns of PA because the reference period to measure the PA is previous to their administration [40,48]. However, a large number of available questionnaires for the paediatric population and the lack of high-quality studies make it difficult to establish recommendations about the best questionnaire available [49–56].

In response to the global need for comparable measures of PA within and between countries, an initiative started in 1996 to develop a compilation of valid questionnaires that culminated in the integration of an International Consensus Group that designed a total of eight versions of the International Physical Activity Questionnaire (IPAQ). The aim of this group was to consolidate a line of indirect measures that could monitor and guide the development of policies focused on PA related to health. Since then, this set of questionnaires has been expanded with the creation of versions focused on other populations and with adaptations for different countries and cultures [57,58]. Therefore, the interest continues with the study of the instruments of the IPAQ group, specifically focusing on the questionnaires available for preschoolers and children. There are various adapted versions of IPAQ questionnaires for different ranges of age in paediatrics. On the one hand, the version for children from 8 to 14 years old has been published and named Physical Activity Questionnaire-Children (PAQ-C) [59,60]. Besides, the English version of the questionnaire for children under 5 years old, known as the Early Years Physical Activity Questionnaire (EY-PAQ), is to be answered by parents [61]. Nevertheless, there is no adapted version for children aged 5–7 years old. Therefore, this study aimed to design the Physical Activity Questionnaire-Young Children (PAQ-YC) to evaluate the PA level in children aged 5–7 years old and evaluate its content validity in terms of face validity, relevance, comprehensiveness and comprehensibility.

2. Materials and Methods

The study of the development of the Physical Activity Questionnaire-Young Children (PAQ-YC) was conducted following the directives of design and analysis of the Consensus-based Standards for the Selection of Health Measurement Instruments guideline (COSMIN) [62].

2.1. Description of the Construct

PAQ-YC aims to measure the total level of PA done by a child aged 5–7 years old in a typical week through the school term. It includes PA at school (Physical Education or other similar activities and break times), transport and leisure time (after-school activities, active games at home and indoor equipment, and outdoor sports) during the last 7 days, assuming that they can be understood as a representation of a typical week unless an extraordinary event could have occurred and could have prevented the fulfilment of the daily activities. Additionally, two questions related to sedentary behaviour during leisure time have been included.

The PAQ-YC design was adjusted to its supposed application with discriminative purposes, with the intention that the questionnaire results helped to identify active enough and inactive subjects, following the international recommendations on PA [63]. On its development, the time of administration no longer than 15 min was considered, and the answers given by parents, children and teachers were unified in a unique version.

2.2. Selection of the Items to Obtain the Final Version

The PAQ-YC design was based on the conceptual framework defined by Pettee Gabriel KK et al. [41], with its adaptations for the paediatric population. The method used to choose the items was preceded by a complete revision of the literature that allowed a compilation of questions susceptible to be included in the final version. The development of the PAQ-YC took as a reference the published versions for different ranges of age and the indications of the literature regarding the type of PA usually done by children from 5 to 7 years old. The PAQ-C version validated for the Spanish population by Manchola-González et al. for children ranged 8–14 years old [64], and the EY-PAQ version validated by Bingham et al. for children less than five years old [61] were used as a model. The final title for the PAQ-YC was chosen to emphasise the targeted range of age.

Later, the answer options for each item were described. For questions about leisure time and transport, a grid format including predefined time ranges was used. For the items referred to school, a 3-item scale, in which the respondent had to mark the closest to the usual conduct, was defined. Regarding its structure, there were two parts: questions where parents could answer by themselves (part 1) and questions that needed the participation of the child or teacher to confirm the answers given (part 2).

A Delphi survey was created to evaluate the content validity. The proposal of collaboration was sent to potential experts related to PA in paediatrics. To evaluate the suitability of the selected experts, the knowledge coefficient (Kc) was used, obtaining their expertise level by calculating the weighted average of the punctuation obtained in each of the items shown in Table 1. Kc was calculated regarding the information that the expert showed about the topic of research. Experts that did not reach the demandable critical level of 0.8 were excluded [65].

Table 1. Knowledge coefficient (Kc) ponderation.

Evaluation Items	Ponderation
Knowledge of PA in educational centres in Spain	0.20
Knowledge of PA done by children at school period	0.20
Practical experience in PA in school stage	0.25
Knowledge about recommendations of PA in school stage	0.25
Abilities in using instruments to evaluate the level of PA	0.10

Afterward, the questionnaire consultation rounds were done with the members of the panel of experts included. An online survey was elaborated to generate a debate that allowed ordered feedback. The consultation rounds suggested an iterative process and were repeated as often as necessary until an agreement was reached. The identity of the participants was anonymous.

The experts were asked for expressing their level of agreement with each of the items, as well as the general characteristics of the questionnaire to evaluate its face validity. The level of agreement was expressed in a rating scale from 1 to 5, in which 1 was the lowest level possible “strongly disagree”, 2 for “disagree”, 3 for “indifference”, 4 for “agree”, and 5 for “strongly agree”. Moreover, a separate section for comments was enabled to allow the participants to manifest specific improvement suggestions.

For each consultation round, the median (Me) and quartiles 1 (Q1) and 3 (Q3) of the answers obtained were calculated. It was established as a consensus criterion that the relative interquartile range ($RIR = (Q3 - Q1/Me) \times 100$) and/or the coefficient of variation ($CV = (SD/Me) \times 100$) were less or equal to 20% [66].

Once the consensual version by the experts was obtained, a process of cognitive interviews in a representative sample of parents was performed. This study received the approval of the Ethical Committee of the Universitat Internacional de Catalunya (Code: FIS-2018-06). First, they were given the paper questionnaire, and they were asked to self-report it. Then, a semi-structured interview was done, where they were asked about the relevance, comprehensiveness and comprehensibility of each of the items and the general

characteristics. Lastly, the spontaneously mentioned issues were commented. A verbal report was the approach used. Notes were written during the process.

3. Results

The first version of the questionnaire was evaluated by the Delphi survey. The final panel was comprised of 13 experts (7 men, 6 women, mean age 38.85 years (11.43); $K_c = 8.67$ (0.7)) on the first round and 11 experts (5 men, 6 women, mean age 39.91 years (11.99); $K_c = 8.50$ (0.61)) on the second round. In the first round, 2 Physiotherapists, 2 graduated in Sports Science, 2 Physiotherapists and Graduated in Sports Science, 4 Primary School teachers, 1 Preschooler teacher and 1 Preschooler teacher and Psychologist participated. In the second round, two of the Primary School teachers that participated in the first round did not participate in the second one.

Two rounds were needed to reach the general agreement. In the first round, the experts answered “strongly agree” or “agree” on items 1–5 of the first part, on the representativity of a usual week item and the general characteristics of the questionnaire. In the second round, experts answered “strongly agree” or “agree” on items 1–5 of part 1, on items 9 and 10 of part 2, on the item about representativity of a usual week and on the general characteristics of the questionnaire. It is essential to mention that on items 6 and 8, only one expert did not answer “strongly agree” or “agree”, whilst 2 experts showed disagreement on item 7 compared to the general opinion (Table 2).

Table 2. Response rates for each of the items and general characteristics.

Variable	Round 1					Round 2				
	1	2	3	4	5	1	2	3	4	5
% of Agreement										
Item 1				38	62					100
Item 2				31	69				36	64
Item 3				23	77					100
Item 4				23	77				9	91
Item 5				15	85				9	91
Item 6		8	8	15	69			9	27	64
Item 7		8	15	31	46	18			27	55
Item 8		8	31	38	23			10	45	45
Item 9			31	31	38				18	82
Item 10			24	38	38				36	64
Representativity of a usual week				23	77				9	91
Clarity of the introductory content				31	69				18	82
Comprehension easiness				38	62				9	91
Clarity and simplicity of the format				54	46				36	64
Division in two parts				23	77				9	91
Number of questions				23	77				9	91
Content appropriateness				31	69				18	82

1 “strongly disagree”, 2 “disagree”, 3 “indifference”, 4 “agree”, 5 “strongly agree”, dashed line: separation between part 1 and 2.

The second part questions and question 7 of part 1 were motives of discrepancies, showing a RIR and CV higher than 20%. However, a general agreement could be considered due to the accomplishment of at least one of the established criteria (item 6: RIR = 20.00 and CV = 15.45; item 7: RIR = 20.00 and CV = 38.73; item 8: RIR = 25.00 and CV = 15.45). Table 3 shows the degree of the agreement for each of the items and general characteristics in the consulting rounds.

Once the consented version by experts was done, the questionnaire was submitted to a process of cognitive interviews in a total of 5 heterogeneous families regarding sex (3 boys, 2 girls, 5 mothers, 2 fathers), age (children 5.8 (0.84) years old; parents 36 (5) years old), parent’s level of education [67] (1 Lower Secondary School, 1 Upper Secondary School, 1 Professional Training, 4 University education or equivalent) and demographic environment (2 in a village between 5.000–10.000 inhabitants, 1 in a town between 10.000–20.000 inhabi-

tants and 4 in a city of more than 100.000 inhabitants). It is important to mention that both parents were interviewed separately in two families, so there were 7 adults interviewed.

The responsible investigator (MA) was familiar with the use and validation of questionnaires in paediatrics [68,69]. Moreover, she was trained in developing PA questionnaires addressed to the general infantile population. The interviews took between 30 and 40 min. Improvement suggestions were not very relevant and were mainly related to explanations in the definition of the questions. No specific approach was used to evaluate the collected information because the suggestions made by the families were simple and did not specify the need for making essential changes. Moreover, as the modifications were not significant, it was considered that there was not a lack of essential aspects on the construct, and the developing process of the PAQ-YC was finished. The final version of the PAQ-YC is accessible in the supplementary material.

Table 3. Degree of the agreement for each of the items and general characteristics in the consulting rounds.

	Variable	Min	Max	Q1	Me	Q 3	RIR	SD	Mean	CV
Round 1	Item 1	4	5	4	5	5	20	0.51	4.62	10.97
	Item 2	4	5	4.75	5	5	5	0.48	4.69	10.24
	Item 3	4	5	5	5	5	0	0.44	4.77	9.19
	Item 4	4	5	5	5	5	0	0.44	4.77	9.19
	Item 5	4	5	5	5	5	0	0.38	4.85	7.75
	Item 6	2	5	4	5	5	20	0.97	4.46	21.68
	Item 7	2	5	3.75	4.5	5	27.78	0.99	4.15	23.76
	Item 8	2	5	3	4	4	25	0.93	3.77	24.59
	Item 9	3	5	3	4	5	50	0.86	4.08	21.15
	Item 10	3	5	3.75	4	5	31.25	0.8	4.15	19.27
	Representativity of a usual week	4	5	5	5	5	0	0.44	4.77	9.19
	Clarity of the introductory content	4	5	4.75	5	5	5	0.48	4.69	10.24
	Comprehension easiness	4	5	4	5	5	20	0.51	4.62	10.97
	Clarity and simplicity of the format	4	5	4	4.5	5	22.22	0.52	4.46	11.63
	Division in two parts	4	5	4.75	5	5	5	0.44	4.77	9.19
Number of questions	4	5	5	5	5	0	0.44	4.77	9.19	
Content appropriateness	4	5	4.75	5	5	5	0.5	4.7	10.24	
Round 2	Item 1	5	5	5	5	5	0	0	5	0
	Item 2	4	5	4	5	5	20	0.5	4.64	10.88
	Item 3	5	5	5	5	5	0	0	5	0
	Item 4	4	5	5	5	5	0	0.3	4.91	6.14
	Item 5	4	5	5	5	5	0	0.3	4.91	6.14
	Item 6	3	5	4	5	5	20	0.69	4.55	15.13
	Item 7	1	5	4	5	5	20	1.55	4	38.73
	Item 8	3	5	4	4	5	25	0.67	4.36	15.45
	Item 9	4	5	5	5	5	0	0.4	4.82	8.4
	Item 10	4	5	4	5	5	20	0.5	4.64	10.88
	Representativity of a usual week	4	5	5	5	5	0	0.3	4.91	6.14
	Clarity of the introductory content	4	5	5	5	5	0	0.4	4.82	8.4
	Comprehension easiness	4	5	5	5	5	0	0.3	4.91	6.14
	Clarity and simplicity of the format	4	5	4	5	5	20	0.5	4.64	10.88
	Division in two parts	4	5	5	5	5	0	0.3	4.91	6.14
Number of questions	4	5	5	5	5	0	0.3	4.91	6.14	
Content appropriateness	4	5	5	5	5	0	0.4	4.82	8.4	

Min: minimum, Max: maximum, Q1: quartile 1, Me: mean, Q3: quartile 3, RIR: Relative Interquartile Range, SD: standard deviation, CV: Coefficient of Variation, dashed line: separation between part 1 and 2

4. Discussion

The first version of the PAQ-YC was designed through a selection and reduction of items process and was tested by an online 2-round Delphi survey among experts in physical activity and a representative sample of the target population. The results

showed adequate content validity in terms of face validity, relevance, comprehensibility and comprehensiveness of the items and the clarity of the general characteristics.

Regarding the construct, PAQ-YC pretends to capture the total PA done in all intensity ranges and in all domains. At this juncture, among the available questionnaires for the range of age of interest in this study, the great variety of constructs evaluated stand out. For example, the Netherlands Physical Activity Questionnaire (NPAQ) was designed to evaluate the preferences of activities assuming an association with the levels of PA [70,71], meanwhile the Harro's questionnaire [72,73], the Assessment of Young Children's Activity using Video Technology (ACTIVITY) questionnaire [74] and the South American Youth/Child Cardiovascular and Environment Study Physical Activity (SAYCARE) questionnaire [75] quantify the total PA so that they could be similar to PAQ-YC. On the other side, questionnaires that include at least two years of the age range of the population of PAQ-YC (5–6 years old or 6–7 years old) gather more delimited constructs, focusing on the evaluation of a higher intensity PA. For example, the Children's Leisure Activities Study Survey (CLASS) questionnaire [76], the School Health Action, Planning and Evaluation System (SHAPES) questionnaire [77], the Bringolf-Isler questionnaire [78], the Canadian Health Measures Survey (CHMS) questionnaire [79] or the England Health Survey [80] inform about moderate to vigorous PA. Compared to the questionnaires included in the IPAQ group for paediatric ages, EY-PAQ [61], PAQ-C [59,60] and Physical Activity Questionnaire-Adolescents (PAQ-A) [81] evaluate the PA in the context of moderate to vigorous intensity, unlike PAQ-YC, that measures PA in all the different levels of intensity.

With reference to the recall period, PAQ-YC goes back to the last 7 days. In PA questionnaires aimed for the same or close range of age, recall periods are very different. For example, NPAQ inquires about the PA of the previous 6 months [70,71]; EY-PAQ gathers activities of the last month [61]; the SAYCARE questionnaire [75], the England Survey of Health of Physical Activity [80]; SHAPES questionnaire [77], PAQ-C [59,60] and PAQ-A [81] refer to the last 7 days; Harro's questionnaire [72,73] and ACTIVITY questionnaire [74] evaluate the PA of the previous day, and Bringolf-Isler questionnaire gathers two typical days [78]. Experts point out that the answers depend on recall bias to define activity patterns [40,44,82], especially complex in paediatrics due to the flashing characteristic of PA [37,83]. About the "typical" notion, there is disagreement among experts regarding its adequacy, although some authors consider that the word creates confusion [57,84].

Another critical point in the initial formulation of the items of PAQ-YC was the decision of quantifying PA in terms of type, duration, frequency and intensity. Building this part was extremely difficult due to the sporadic behaviour of children's PA, which makes this activity challenging to remember, quantify and classify [37].

For items 1–6, an approach based on categories of activity with an answers table that pretends to gather detailed information about frequency and duration was used. Concerning the type of activity, offering closed lists was ruled out because the variety of PA done by children makes it difficult to have a moderate-extension inventory without dismissing any significant action [85]. Nevertheless, examples of activities included in every item were given to facilitate remembrance and clarify the type of activities asked in each question [86]. About the format of other questionnaires for the same or similar range of age, some use the same approach as a checklist, like NPAQ [70,71], CLASS [76], Bringolf-Isler questionnaire [78], or EY-PAQ [61], while others classify PA in categories, as Harro's questionnaire [72,73], ACTIVITY [74], SAYCARE [75], CHMS [79] or SHAPES [77]. Finally, PAQ-C [64] and PAQ-A [87] combine both formats, as they show items for activity categories and a 22-activity checklist. Regarding questionnaires that classify PA in categories and PAQ-YC, it is usual that they clarify which type of activities they are referring to in each item [72,73,79]. In PAQ-C [59,60] and PAQ-A [81], only some of the questions give concrete examples of activities.

The duration was expressed in reduced intervals in part 1, while the frequency was gathered for every day of the week. For part 2, options of answers were based on a closed

scale. Activities during the school period are developed as usual behaviour and have little variability, so the opportunity for PA remains constant and can be estimated [88–90]. Questionnaires for the same or similar range of age differ from PAQ-YC on the format and detailed expression to gather duration and frequency. For example, the original version of NPAQ does not consider duration nor frequency [70], whilst it includes one item about time spent watching television on its modified version [71]; Harro's questionnaire [72,73], SAYCARE [75] and CLASS [76] gather the duration and frequency for every activity; ACTIVITY divides the day into 6 periods, and registers the activities done in each period [74]; SHAPES indicate the number of hours and the increasing of 15 min for every day of the week on its questions [77], while CHMS [79] and Bringolf-Isler questionnaire [78] only ask for information about the duration of each activity, with similar intervals as PAQ-YC. EY-PAQ [61] is also like the first part of PAQ-YC. Nevertheless, the manner of quantifying frequency differs sustainably from it, as it asks for the number of days of the week that each activity is done, while PAQ-YC is supposed to be filled every day. It is important to mention that PAQ-C [59,60] and PAQ-A [81] do not ask for information of duration in any case. In this context, the graph format of item 9 of PAQ-C or item 8 of PAQ-A is like the design of the options of answers in the first part of PAQ-YC.

Regarding the intensity, the average of the energetic cost for every activity category was calculated following the Youth Energy Expenditure Compendium [91]. For the questionnaires commented on previously, the way that intensity is evaluated differs depending on the format. For example, NPAQ does not gather the intensity of the activities on the checklist [70,71]; Harro's questionnaire splits activities from low to moderate intensity and activities from moderate to vigorous intensity [72,73]; SHAPES [77] and SAYCARE [75] define intensity in relative terms; ACTIVITY uses videoclips from 3 to 5 s that show levels of intensity [74]; CLASS [76] classifies the activities of the checklist in moderate or high intensity allocating MET values from the Adult Compendium [92], unlike PAQ-YC, that uses the references from the Youth Energy Expenditure Compendium [93]; while in Bringolf-Isler [78] and CHMS [79] PA is included in the moderate-to-vigorous category. Gathering intensity in relative terms was excluded because the effort perception depends on individual characteristics [94]. It was also considered to be inappropriate to express intensity only in terms of effort, such as sweating, increasing cardiac frequency or dyspnoea [95,96]. Therefore, and even though it is admitted that the measurement error always occurs, scientific literature shows that it can be minimised by describing intensity in absolute terms [47,84].

Once the first version of the PAQ-YC was designed, the content validity study began through a double-consultation process [97,98]. The process of consulting experts and representative samples are relevant when creating any questionnaire but are more important when the instrument is based on a formative model, as PA questionnaires [99].

The process of selection of the experts represents the focal point of the survey. In our study, we decided to establish different items to calculate a median average according to the importance (Kc). We did not consider any external criterion because, being a heterogeneous group, inconsistency could occur. Other authors, such as Cabero-Almenara suggests leaning on the expert self-report, as well as on the sources that allow increasing the answer given [65,100]. On the other hand, the COSMIN group, in its international Delphi study, establish as an expertise criterion to have at least 5 publications on PubMed about the area of interest [98]. Regarding the number of experts, it was considered a number between 10 and 15, provided that the necessary heterogeneity could be ensured to approach the aim of the consulting through the composition of the group. For example, Ruiz-Olabuénaga and Ispuzua suggest a number between 10 and 30 [101]; García and Fernández between 15 and 25 [102], while others suggest that for operativity issues, it is not recommended to have more than 50 experts [103]. Next, the agreed version of the PAQ-YC with the panel of experts was submitted to a cognitive interviews process in a representative sample of the population. Due to the lack of agreement, the principle of saturation was prioritised [86,104] over the number of interviewees, even though it was

decided to recruit at least 7 participants. Some authors suggest that between 7 and 10 interviews are enough [105–107], while others defend that a large sample is needed [108]. It is recommended to establish it depending on the variability of the population characteristics and the key aspects of the construct [86].

Regarding the quantitative information collection method, the semi-structured interview was used and followed the same structure as on the Delphi survey to ensure a comprehensive and ordered approach but allowing certain flexibility to adapt the interview to the families' concerns. The election of the semi-structured interview was justified by the need of exploring the issue deeply over focal groups, which main strength is their capacity to identify a variety of inter-individual experiences and perspectives. In addition, it has been documented that some individuals may be reluctant to express their point of view in front of a group so that some information could be lost [86].

The approach used to interrogate participants was a verbal report. While it is true that more approaches are valid, some authors consider that a verbal report is the most appropriate way to evaluate the participant's familiarity with the issue and the terminology used. Furthermore, it has been recognised that the interviewer may obtain certain information that otherwise the participant would not give [86,107,109].

Although some authors recommend transcribing records of the interviews to collect data, on the cognitive validation study of PAQ-YC, the decision was only to take notes. This decision was made based on three reasons. First, one family expressed inconvenience in being recorded during the interview. For that, the COSMIN guideline recommends taking notes when participants do not feel comfortable when recording [97]. Secondly, literal transcription methods are designed to document more complex interviews, usually performed in focus groups [110]. Nevertheless, interviews performed for PAQ-YC were made individually, so the process was more straightforward, and the notes taken were enough. Thirdly, the recording could bring problems related to data protection policy [86].

Furthermore, to analyse data collected during interviews, some authors recommend using complex methods designed for procedures with a larger number of participants and a larger flow of information [110]. Instead of this, in agreement with Brod et al. [86], data analysis was done iteratively, beginning as soon as first opinions about the questionnaire were obtained.

Once the interviewing process was finished, the content validity study of the PAQ-YC was terminated [99]. As the contributions made by the population were not very relevant, the commented issues were included without re-evaluating the final version [84]. With regard to the lack of major contributions to the questionnaire, we think it was due to the exhaustive phase of obtaining concepts. COSMIN recognises that the risk in avoiding important aspects is low when the previous phase to the item reduction and experts' consulting is done carefully [84].

5. Limitations

The main limitation of this study is related to the difficulty of designing questionnaires for the 5 to 7-year age group. It includes a population between Pre-Primary Education and Primary Education, that combines times when parents are not in charge of their children, but children do not have enough resources to answer the questionnaire by themselves [37,111,112]. A parent-reported questionnaire was opted for in PAQ-YC, pretending to abridge the provided information by children or teachers at school. This decision was taken based on the difficulty of parents in reporting the PA done when they are not there [113,114] and on that the expert's recommendations about self-reporting say that it should be only applied beyond 10 years old [49,114–117].

6. Conclusions

The findings of development and content validity of the PAQ-YC show that it is adequate in terms of relevance, comprehensiveness and comprehensibility of the items.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/healthcare9060655/s1>.

Author Contributions: M.A.-B.: conception, design, data analysis and drafting; M.G.-F.: conception and data analysis; F.R.-N.: conception and drafting; A.O.-M.: design and drafting; A.M.-R.: design and data collect; C.B.-C.: design and drafting. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethical Committee of the Universitat Internacional de Catalunya (Code: FIS-2018-06).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Remmers, T.; Sleddens, E.F.C.; Gubbels, J.S.; de Vries, S.I.; Mommers, M.; Penders, J.; Kremers, S.P.J.; Thijs, C. Relationship between physical activity and the development of body mass index in children. *Med. Sci. Sports Exerc.* **2014**, *46*, 177–184. [\[CrossRef\]](#)
2. Jago, R.; Baranowski, T.; Baranowski, J.C.; Thompson, D.; Greaves, A.K. BMI from 3–6 y of age is predicted by TV viewing and physical activity, not diet. *Int. J. Obes.* **2005**, *29*, 557–564. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Moore, L.L.; Gao, D.; Bradlee, M.L.; Cupples, L.A.; Sundarajan-Ramamurti, A.; Proctor, M.H.; Hood, M.Y.; Singer, M.R.; Ellison, R.C. Does early physical activity predict body fat change throughout childhood? *Prev. Med.* **2003**, *37*, 10–17. [\[CrossRef\]](#)
4. Jackowski, S.A.; Baxter-Jones, A.D.G.; Gruodyte-Raciene, R.; Kontulainen, S.A.; Erlandson, M.C. A longitudinal study of bone area, content, density, and strength development at the radius and tibia in children 4–12 years of age exposed to recreational gymnastics. *Osteoporos. Int.* **2015**, *26*, 1677–1690. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Janz, K.F.; Letuchy, E.M.; Gilmore, J.M.E.; Burns, T.L.; Torner, J.C.; Willing, M.C.; Levy, S.M. Early physical activity provides sustained bone health benefits later in childhood. *Med. Sci. Sports Exerc.* **2010**, *42*, 1072–1078. [\[CrossRef\]](#)
6. Janz, K.F.; Gilmore, J.M.E.; Levy, S.M.; Letuchy, E.M.; Burns, T.L.; Beck, T.J. Physical activity and femoral neck bone strength during childhood: The Iowa bone development study. *Bone* **2007**, *41*, 216–222. [\[CrossRef\]](#)
7. Beets, M.W.; Beighle, A.; Erwin, H.E.; Huberty, J.L. After-school program impact on physical activity and fitness: A meta-analysis. *Am. J. Prev. Med.* **2009**, *36*, 527–537. [\[CrossRef\]](#)
8. Saavedra, J.M.; Escalante, Y.; Garcia-Hermoso, A. Improvement of aerobic fitness in obese children: A meta-analysis. *Pediatr. Obes.* **2011**, *6*, 169–177. [\[CrossRef\]](#)
9. Larouche, R.; Saunders, T.J.; Faulkner, G.E.J.; Colley, R.; Tremblay, M. Associations between active school transport and physical activity, body composition, and cardiovascular fitness: A systematic review of 68 studies. *J. Phys. Act. Health* **2014**, *11*, 206–227. [\[CrossRef\]](#)
10. Dobbins, M.; Husson, H.; DeCorby, K.; LaRocca, R.L. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. *Cochrane Database Syst. Rev.* **2013**, *2013*, CD007651. [\[CrossRef\]](#)
11. Clark, E.J. Does the type of intervention method really matter for combating childhood obesity? A systematic review and meta-analysis. *J. Sports Med. Phys. Fit.* **2014**, *55*, 1524–1543.
12. Malina, R.M. Weight training in youth-growth, maturation, and safety: An evidence-based review. *Clin. J. Sport Med.* **2006**, *16*, 478–487. [\[CrossRef\]](#)
13. Vasconcellos, F.; Seabra, A.; Katzmarzyk, P.; Kraemer-Aguiar, L.G.; Bouskela, E.; Farinatti, P. Physical activity in overweight and obese adolescents: Systematic review of the effects on physical fitness components and cardiovascular risk factors. *Sports Med.* **2014**, *44*, 1139–1152. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Millard-Stafford, M.; Becasen, J.S.; Beets, M.W.; Nihiser, A.J.; Lee, S.M.; Fulton, J.E. Is physical fitness associated with health in overweight and obese youth? A Systematic Review. *Kinesiol. Rev.* **2013**, *2*, 233–247. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Faigenbaum, A.D.; McFarland, J.E.; Johnson, L.; Kang, J.; Bloom, J.; Ratamess, N.A.; Hoffman, J.R. Preliminary evaluation of an after-school resistance training program for improving physical fitness in middle school-age boys. *Percept. Mot. Ski.* **2007**, *104*, 407–415. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Wilks, D.C.; Sharp, S.J.; Ekelund, U.; Thompson, S.G.; Mander, A.P.; Turner, R.M.; Jebb, S.A.; Lindroos, A.K. Objectively measured physical activity and fat mass in children: A bias-adjusted meta-analysis of prospective studies. *PLoS ONE* **2011**, *6*, e17205. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Nogueira, R.C.; Weeks, B.K.; Beck, B.R. Exercise to improve pediatric bone and fat. *Med. Sci. Sports Exerc.* **2014**, *46*, 610–621. [\[CrossRef\]](#)

18. Ramires, V.V.; Dumith, S.C.; Gonçalves, H. Longitudinal association between physical activity and body fat during adolescence: A systematic review. *J. Phys. Act. Health* **2015**, *12*, 1344–1358. [[CrossRef](#)] [[PubMed](#)]
19. Janssen, I.; LeBlanc, A.G. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* **2010**, *7*, 40. [[CrossRef](#)]
20. Tan, V.P.S.; Macdonald, H.; Kim, S.; Nettlefold, L.; Gabel, L.; Ashe, M.C.; McKay, A.H. Influence of physical activity on bone strength in children and adolescents: A Systematic review and narrative synthesis. *J. Bone Miner. Res.* **2014**, *29*, 2161–2181. [[CrossRef](#)]
21. Weaver, C.M.; Gordon, C.M.; Janz, K.F.; Kalkwarf, H.J.; Lappe, J.M.; Lewis, R.; O’Karma, M.; Wallace, T.C.; Zemel, B.S. The National Osteoporosis Foundation’s position statement on peak bone mass development and lifestyle factors: A systematic review and implementation recommendations. *Osteoporos. Int.* **2016**, *27*, 1281–1386. [[CrossRef](#)] [[PubMed](#)]
22. Kelley, G.A.; Kelley, K.S. Aerobic exercise and lipids and lipoproteins in children and adolescents: A meta-analysis of randomized controlled trials. *Atherosclerosis* **2007**, *191*, 447–453. [[CrossRef](#)]
23. Fedewa, M.V.; Gist, N.H.; Evans, E.M.; Dishman, R.K. Exercise and insulin resistance in youth: A meta-analysis. *Pediatrics* **2013**, *133*, e163–e174. [[CrossRef](#)] [[PubMed](#)]
24. Ahn, S.; Fedewa, A.L. A Meta-analysis of the relationship between children’s physical activity and mental health. *J. Pediatr. Psychol.* **2011**, *36*, 385–397. [[CrossRef](#)]
25. Biddle, S.J.H.; Asare, M. Physical activity and mental health in children and adolescents: A review of reviews. *Br. J. Sports Med.* **2011**, *45*, 886–895. [[CrossRef](#)]
26. Pastor, Y. Testing direct and indirect effects of sports participation on perceived health in Spanish adolescents between 15 and 18 years of age. *J. Adolesc.* **2003**, *26*, 717–730. [[CrossRef](#)]
27. Parfitt, G.; Eston, R.G. The relationship between children’s habitual activity level and psychological well-being. *Acta Paediatr.* **2007**, *94*, 1791–1797. [[CrossRef](#)]
28. Ruiz-Hermosa, A.; Álvarez-Bueno, C.; Cervero-Redondo, I.; Martínez-Vizcaino, V.; Redondo-Tébar, A.; Sánchez-López, M. Active commuting to and from school, cognitive performance, and academic achievement in children and adolescents: A systematic review and meta-analysis of observational studies. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1839. [[CrossRef](#)] [[PubMed](#)]
29. Álvarez-Bueno, C.; Pesce, C.; Cervero-Redondo, I.; Sánchez-López, M.; Martínez-Hortelano, J.A.; Martínez-Vizcaino, V. The effect of physical activity interventions on children’s cognition and metacognition: A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **2017**, *56*, 729–738. [[CrossRef](#)] [[PubMed](#)]
30. Sánchez-López, M.; Cervero-Redondo, I.; Alvarez-Bueno, C.; Ruiz-Hermosa, A.; Pozuelo-Carrascosa, D.P.; Diez-Fernández, A.; Del Campo, D.G.-D.; Pardo-Guijarro, M.J.; Martínez-Vizcaino, V. Impact of a multicomponent physical activity intervention on cognitive performance: The MOVI-KIDS study. *Scand. J. Med. Sci. Sports* **2019**, *29*, 766–775. [[CrossRef](#)] [[PubMed](#)]
31. Carson, V.; Hunter, S.; Kuzik, N.; Gray, C.; Poitras, V.J.; Chaput, J.-P.; Saunders, T.J.; Katzmarzyk, P.; Okely, A.; Gorber, S.C.; et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: An update. *Appl. Physiol. Nutr. Metab.* **2016**, *41*, S240–S265. [[CrossRef](#)]
32. Tremblay, M.S.; LeBlanc, A.G.; Kho, M.E.; Saunders, T.J.; Larouche, R.; Colley, R.C.; Goldfield, G.; Gorber, S.C. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. *Int. J. Behav. Nutr. Phys. Act.* **2011**, *8*, 98. [[CrossRef](#)]
33. Cliff, D.P.; Hesketh, K.; Vella, S.A.; Hinkley, T.; Tsiros, M.D.; Ridgers, N.D.; Carver, A.; Veitch, J.; Parrish, A.-M.; Hardy, L.L.; et al. Objectively measured sedentary behaviour and health and development in children and adolescents: Systematic review and meta-analysis. *Obes. Rev.* **2016**, *17*, 330–344. [[CrossRef](#)]
34. Herting, M.M.; Chu, X. Exercise, cognition, and the adolescent brain. *Birth Defects Res.* **2017**, *109*, 1672–1679. [[CrossRef](#)]
35. Plasqui, G.; Westerterp, K.R. Physical activity assessment with accelerometers: An evaluation against doubly labeled water**. *Obesity* **2007**, *15*, 2371–2379. [[CrossRef](#)]
36. Haugen, H.A.; Chan, L.-N.; Li, F. Indirect calorimetry: A practical guide for clinicians. *Nutr. Clin. Pract.* **2007**, *22*, 377–388. [[CrossRef](#)] [[PubMed](#)]
37. Sirard, J.R.; Pate, R.R. Physical Activity assessment in children and adolescents. *Sports Med.* **2001**, *31*, 439–454. [[CrossRef](#)] [[PubMed](#)]
38. Pals, M.A.; Peeters, P.H.; Kemper, H.C.; Grobbee, D.E. Methodological aspects of physical activity assessment in epidemiological studies. *Eur. J. Epidemiol.* **1998**, *14*, 63–70. [[CrossRef](#)] [[PubMed](#)]
39. Patterson, P. Reliability, validity, and methodological response to the assessment of physical activity via self-report. *Res. Q. Exerc. Sport* **2000**, *71*, 15–20. [[CrossRef](#)] [[PubMed](#)]
40. LaMonte, M.J.; Ainsworth, B.E. Quantifying energy expenditure and physical activity in the context of dose response. *Med. Sci. Sports Exerc.* **2001**, *33*, S370–S378. [[CrossRef](#)]
41. Gabriel, K.K.P.; Morrow, J.R.; Woolsey, A.-L.T. Framework for physical activity as a complex and multidimensional behavior. *J. Phys. Act. Health* **2012**, *9*, S11–S18. [[CrossRef](#)] [[PubMed](#)]
42. Kesaniemi, Y.K.; Danforth, E.; Jensen, M.D.; Kopelman, P.G.; Lefèbvre, P.; Reeder, A.B. Dose-response issues concerning physical activity and health: An evidence-based symposium. *Med. Sci. Sports Exerc.* **2001**, *33*, S351–S358. [[CrossRef](#)]
43. Terwee, C.B.; Mokkink, L.B.; Van Poppel, M.; Chinapaw, M.J.M.; Van Mechelen, W.; De Vet, H.C.W. Qualitative attributes and measurement properties of physical activity questionnaires. *Sports Med.* **2010**, *40*, 525–537. [[CrossRef](#)] [[PubMed](#)]

44. Warren, J.M.; Ekelund, U.; Besson, H.; Mezzani, A.; Geladas, N.; Vanhees, L. Assessment of physical activity—A review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur. J. Cardiovasc. Prev. Rehabilitation* **2010**, *17*, 127–139. [CrossRef] [PubMed]
45. Strath, S.J.; Kaminsky, L.A.; Ainsworth, B.E.; Ekelund, U.; Freedson, P.S.; Gary, R.A.; Richardson, C.R.; Smith, D.T.; Swartz, A.M. Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart Association. *Circulation* **2013**, *128*, 2259–2279. [CrossRef]
46. Ainsworth, B.; Cahalin, L.; Buman, M.; Ross, R. The current state of physical activity assessment tools. *Prog. Cardiovasc. Dis.* **2015**, *57*, 387–395. [CrossRef]
47. Lagerros, Y.T.; Lagiou, P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *Eur. J. Epidemiol.* **2007**, *22*, 353–362. [CrossRef]
48. Lagerros, Y.T. Physical activity—The more we measure, the more we know how to measure. *Eur. J. Epidemiol.* **2009**, *24*, 119–122. [CrossRef]
49. Kohl, H.W.; Fulton, J.E.; Caspersen, C.J. Assessment of physical activity among children and adolescents: A review and synthesis. *Prev. Med.* **2000**, *31*, S54–S76. [CrossRef]
50. Tessier, S.; Vuillemin, A.; Briançon, S. Review of physical activity questionnaires validated for children and adolescents. *Sci. Sports* **2008**, *23*, 118–125. [CrossRef]
51. Chinapaw, M.J.; Mokkink, L.B.; Van Poppel, M.N.M.; Van Mechelen, W.; Terwee, C.B.; Chinapaw, M.J.M. Physical activity questionnaires for youth. *Sports Med.* **2010**, *40*, 539–563. [CrossRef]
52. Foley, L.; Maddison, R.; Olds, T.; Ridley, K. Self-report use-of-time tools for the assessment of physical activity and sedentary behaviour in young people: Systematic review. *Obes. Rev.* **2012**, *13*, 711–722. [CrossRef]
53. Cancela, J.M.; Ayán, C.; Castro, A. An evaluation of questionnaires assessing physical activity levels in youth populations. *J. Child Health Care* **2013**, *17*, 274–293. [CrossRef] [PubMed]
54. Martínez-Lemos, R.; Pérez, C.A.; Lastra, A.S.; Carral, J.M.C.; Sánchez, R.V. Physical activity questionnaires for Spanish children and adolescents: A systematic review. *An. Sist. Sanit. Navar.* **2016**, *39*, 417–428. [CrossRef]
55. Hidding, L.M.; Chinapaw, M.J.M.; Van Poppel, M.N.M.; Mokkink, L.B.; Altenburg, T.M. An updated systematic review of childhood physical activity questionnaires. *Sports Med.* **2018**, *48*, 2797–2842. [CrossRef] [PubMed]
56. Yang, X.; Zhai, Y.; Si, X.; Zhao, W.H. Validity and reliability of physical activity questionnaires in children and adolescents: A meta-analysis. *Zhonghua Yu Fang Yi Xue Za Zhi* **2020**, *54*, 546–554. [PubMed]
57. Craig, C.L.; Marshall, A.L.; Sjöström, M.; Bauman, A.E.; Booth, M.L.; Ainsworth, B.E.; Pratt, M.; Ekelund, U.; Yngve, A.; Sallis, J.F.; et al. International physical activity questionnaire: 12-country reliability and Validity. *Med. Sci. Sports Exerc.* **2003**, *35*, 1381–1395. [CrossRef]
58. International Physical Activity Questionnaire. Available online: <https://sites.google.com/site/theipaq/> (accessed on 2 March 2020).
59. Kowalski, K.C.; Crocker, P.R.; Faulkner, R.A. Validation of the physical activity questionnaire for older children. *Pediatr. Exerc. Sci.* **1997**, *9*, 174–186. [CrossRef]
60. Crocker, P.R.E.; Bailey, D.A.; Faulkner, R.A.; Kowalski, K.C.; McGrath, R. Measuring general levels of physical activity: Preliminary evidence for the Physical Activity Questionnaire for Older Children. *Med. Sci. Sports Exerc.* **1997**, *29*, 1344–1349. [CrossRef]
61. Bingham, D.D.; Collings, P.J.; Cledes, S.A.; Costa, S.; Santorelli, G.; Griffiths, P.; Barber, S.E. Reliability and validity of the early years physical activity questionnaire (EY-PAQ). *Sports* **2016**, *4*, 30. [CrossRef]
62. Mokkink, L.B.; Prinsen, C.A.C.; Bouter, L.M.; De Vet, H.C.W.; Terwee, C.B. The COnsensus-based Standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. *Braz. J. Phys. Ther.* **2016**, *20*, 105–113. [CrossRef]
63. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Scientific Report 2018. Available online: <https://health.gov/our-work/physical-activity/current-guidelines/scientific-report> (accessed on 2 March 2020).
64. Manchola-González, J.; Bagur-Calafat, C.; Girabent-Farrés, M. Fiabilidad de la versión española del Cuestionario de actividad física PAQ-C / Reliability of the Spanish version of questionnaire of physical Activity PAQ-C. *Rev. Int. Med. Ciencias Actividad Física Deporte* **2017**, *65*, 139–152. [CrossRef]
65. Almenara, J.C.; Moro, A.I. Empleo del método Delphi y su empleo en la investigación en comunicación y educación. *Educat. Rev. Electrónica Tecnol. Educ.* **2014**, *a272*. [CrossRef]
66. Reguant-Álvarez, M.; Torrado-Fonseca, M. El método Delphi. *Rev. Innov. Recer. Educ.* **2016**, *9*, 87–102.
67. Instituto Nacional de Estadística (INE). Clasificación por Nivel de Estudios. Available online: <https://www.ine.es/daco/daco42/discapa/tar3.pdf>. (accessed on 2 March 2020).
68. Farrés, M.G.; Mata, J.F.; Barbosa, M.A.; Calafat, C.B. Traducción y validación al español del módulo neuromuscular de la escala Pediatric Quality of Life Inventory (PedsQL): Evaluación de la calidad de vida autopercebida por niños de 8–18 años con enfermedades neuromusculares y sus padres. *Rev. Neurol.* **2018**, *67*, 425. [CrossRef]
69. Farrés, M.G.; Calafat, C.B.; Barbosa, M.A.; Benito, D.N.-D.; Rincón, A.M.; Mata, J.F. Traducción y validación al español del módulo neuromuscular de la escala Pediatric Quality of Life Inventory (PedsQL): Evaluación de la calidad de vida autopercebida por niños de 5–7 años con enfermedades neuromusculares y por sus padres. *Rev. Neurol.* **2019**, *69*, 442–452. [CrossRef]

70. Montoye, H.; Kemper, H.; Saris, W.; Washburn, R. *Measuring Physical Activity and Energy Expenditure*; Human Kinetics: Champaign, IL, USA, 1996.
71. Janz, K.F.; Broffitt, B.; Levy, S.M. Validation evidence for the Netherlands physical activity questionnaire for young children. *Res. Q. Exerc. Sport* **2005**, *76*, 363–369. [[CrossRef](#)]
72. Harro, M. Validation of a questionnaire to assess physical activity of children ages 4–8 years. *Res. Q. Exerc. Sport* **1997**, *68*, 259–268. [[CrossRef](#)] [[PubMed](#)]
73. Rice, K.R.; Joschtel, B.; Trost, S.G. Validity of family child care providers' proxy reports on children's physical activity. *Child. Obes.* **2013**, *9*, 393–398. [[CrossRef](#)] [[PubMed](#)]
74. Tremblay, M.S.; Inman, J.W.; Willms, J.D. Preliminary evaluation of a video questionnaire to assess activity levels of children. *Med. Sci. Sports Exerc.* **2001**, *33*, 2139–2144. [[CrossRef](#)]
75. Nascimento-Ferreira, M.V.; De Moraes, A.C.F.; Toazza-Oliveira, P.V.; Forjaz, C.L.; Aristizábal, J.C.; Santaliesra-Pasías, A.M.; Lepera, C.; Nascimento-Junior, W.V.; Skapino, E.; Delgado, C.A.; et al. Reliability and validity of a questionnaire for physical activity assessment in South American children and adolescents: The SAYCARE study. *Obesity* **2018**, *26*, S23–S30. [[CrossRef](#)] [[PubMed](#)]
76. Telford, A.; Salmon, J.; Jolley, D.; Crawford, D. Reliability and validity of physical activity questionnaires for children: The children's leisure activities study survey (CLASS). *Pediatr. Exerc. Sci.* **2004**, *16*, 64–78. [[CrossRef](#)]
77. Wong, S.L.; Leatherdale, S.T.; Manske, S.R. Reliability and validity of a school-based physical activity questionnaire. *Med. Sci. Sports Exerc.* **2006**, *38*, 1593–1600. [[CrossRef](#)] [[PubMed](#)]
78. Bringolf-Isler, B.; Mäder, U.; Ruch, N.; Kriemler, S.; Grize, L.; Braun-Fahrlander, C. Measuring and validating physical activity and sedentary behavior comparing a parental questionnaire to accelerometer data and diaries. *Pediatr. Exerc. Sci.* **2012**, *24*, 229–245. [[CrossRef](#)]
79. Colley, R.C.; Wong, S.L.; Garriguet, D.; Janssen, I.; Gorber, S.C.; Tremblay, M.S. Physical activity, sedentary behaviour and sleep in Canadian children: Parent-report versus direct measures and relative associations with health risk. *Health Rep.* **2012**, *23*, A1.
80. Basterfield, L.; Adamson, A.; Parkinson, K.; Maute, U.; Li, P.X.; Reilly, J.J.; The Gateshead Millennium Study Core Team. Surveillance of physical activity in the UK is flawed: Validation of the health survey for England Physical Activity Questionnaire. *Arch. Dis. Child.* **2008**, *93*, 1054–1058. [[CrossRef](#)] [[PubMed](#)]
81. Kowalski, K.C.; Crocker, P.R.; Kowalski, N.P. Convergent validity of the physical activity questionnaire for adolescents. *Pediatr. Exerc. Sci.* **1997**, *9*, 342–352. [[CrossRef](#)]
82. Ainsworth, E.B. How do I measure physical activity in my patients? Questionnaires and objective methods. *Br. J. Sports Med.* **2008**, *43*, 6–9. [[CrossRef](#)]
83. Welk, G.J.; Corbin, C.B.; Dale, D. Measurement issues in the assessment of physical activity in children. *Res. Q. Exerc. Sport* **2000**, *71*, 59–73. [[CrossRef](#)]
84. Altschuler, A.; Picchi, T.; Nelson, M.; Rogers, J.D.; Hart, J.; Sternfeld, B. Physical activity questionnaire comprehension. *Med. Sci. Sports Exerc.* **2009**, *41*, 336–343. [[CrossRef](#)]
85. Sternfeld, B.; Goldman-Rosas, L. A Systematic approach to selecting an appropriate measure of self-reported physical activity or sedentary behavior. *J. Phys. Act. Health* **2012**, *9*, S19–S28. [[CrossRef](#)] [[PubMed](#)]
86. Brod, M.; Tesler, L.E.; Christensen, T.L. Qualitative research and content validity: Developing best practices based on science and experience. *Qual. Life Res.* **2009**, *18*, 1263–1278. [[CrossRef](#)]
87. Martínez-Gómez, D.; Martínez-De-Haro, V.; Pozo, T.; Welk, G.J.; Villagra, A.; Calle, M.E.; Marcos, A.; Veiga, O.L. Fiabilidad y validez del cuestionario de actividad física PAQ-A en adolescentes españoles. *Rev. Española Salud Públ.* **2009**, *83*, 427–439. [[CrossRef](#)] [[PubMed](#)]
88. Ley Orgánica 3/2020, de 29 de Diciembre, Por La Que Se Modifica La Ley Orgánica 2/2006, de 3 de Mayo, de Educación. *Bol. Of. Estado Jefatura Estado* **2020**, *340*, 122868–122953.
89. de España, G. Ley Orgánica 2/2006, de 3 de Mayo, de Educación. Boletín Oficial Del Estado. *Jefatura Estado* **2006**, *106*, 17158–17207.
90. Real Decreto 126/2014, de 28 de Febrero, Por El Que Se Establece el currículo básico de la educación primaria. Ministerio de educación, cultura y deporte. *Bol. Of. Estado* **2014**, *52*, 19349–19420.
91. Butte, N.F.; Watson, K.B.; Ridley, K.; Zakeri, I.F.; McMurray, R.G.; Pfeiffer, K.A.; Crouter, S.E.; Herrmann, S.D.; Bassett, D.R.; Long, A.; et al. A youth compendium of physical activities. *Med. Sci. Sports Exerc.* **2018**, *50*, 246–256. [[CrossRef](#)]
92. Ainsworth, B.E.; Haskell, W.L.; Leon, A.S.; Jacobs, D.R.; Montoye, H.J.; Sallis, J.F.; Paffenbarger, R.S. Compendium of physical activities: Classification of energy costs of human physical activities. *Med. Sci. Sports Exerc.* **1993**, *25*, 71–80. [[CrossRef](#)]
93. Pfeiffer, K.A.; Watson, K.B.; McMurray, R.G.; Bassett, D.R.; Butte, N.F.; Crouter, S.E.; Herrmann, S.D.; Trost, S.G.; Ainsworth, B.E.; Fulton, J.E.; et al. Energy cost expression for a youth compendium of physical activities: Rationale for using age groups. *Pediatr. Exerc. Sci.* **2018**, *30*, 142–149. [[CrossRef](#)]
94. Shephard, R.J. Absolute versus relative intensity of physical activity in a dose-response context. *Med. Sci. Sports Exerc.* **2001**, *33*, S400–S418. [[CrossRef](#)]
95. Shephard, R.J.; Vuillemin, A. Limits to the measurement of habitual physical activity by questionnaires Commentary. *Br. J. Sports Med.* **2003**, *37*, 197–206. [[CrossRef](#)] [[PubMed](#)]
96. Ridley, K.; Ainsworth, E.B.; Olds, T.S. Development of a compendium of energy expenditures for youth. *Int. J. Behav. Nutr. Phys. Act.* **2008**, *5*, 45. [[CrossRef](#)]

97. Terwee, C.B.; Prinsen, C.A.; Chiarotto, A.; Cw De Vet, H.; Bouter, L.M.; Marjan, J.A.; Donald, W.; Patrick, L.; Mokkink, L.B.; Terwee, C.B. COSMIN Standards and criteria for evaluating the content validity of health-related patient-reported outcome measures: A Delphi study. *Qual. Life Res.* **2018**, *27*, 1159–1170. [CrossRef]
98. Mokkink, L.B.; Terwee, C.B.; Patrick, D.L.; Alonso, J.; Stratford, P.W.; Knol, D.L.; Bouter, L.M.; De Vet, H.C.W. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual. Life Res.* **2010**, *19*, 539–549. [CrossRef]
99. De Vet, H.C.W.; Terwee, C.B.; Mokkink, L.B.; Knol, D.L. *Measurement in Medicine*; Cambridge University Press: Cambridge, UK, 2011.
100. Almenara, J.C.; Cejudo, M.; del, C.L. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *Educat. Rev. Electrónica Tecnol. Educ.* **2013**, *7*, 11–22.
101. Ruiz-Olebuénaga, J.; Ispizua-Uribarri, M. *La Descodificación de La Vida Cotidiana*; Deusto University: Bizkaia, Spain, 1998.
102. García-Abreu, L.; Fernández-García, S.J. Procedimiento de aplicación del trabajo creativo en grupo de expertos. *Ing. Energética* **2011**, *29*, 46–50.
103. Landeta, J. *El Método Delphi: Una Técnica de Previsión Para La Incertidumbre*; Editorial Ariel: Barcelona, Spain, 1999.
104. Kerr, C.; Nixon, A.; Wild, D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Rev. Pharm. Outcomes Res.* **2010**, *10*, 269–281. [CrossRef] [PubMed]
105. Cecilia, M.; Prinsen, A.C.; Donald, L.; Jordi, P.; Lex, A.; Bouter, M.; Mokkink, C.L. COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments. 2019. Available online: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf (accessed on 2 March 2020).
106. Leidy, N.K.; Vernon, M. Perspectives on Patient-reported outcomes. *PharmacoEconomics* **2008**, *26*, 363–370. [CrossRef] [PubMed]
107. Willis, G. *Cognitive Interviewing. A Tool for Improving Questionnaire Design*; Thousand, O., Ed.; Sage Publications: Thousand Oaks, CA, USA, 2005.
108. Blair, J.; Conrad, F. Sample size for COGNITIVE interview pretesting. *Public Opin. Q.* **2011**, *75*, 636–658. [CrossRef]
109. Beatty, P.C.; Willis, G.B. Research Synthesis: The practice of cognitive interviewing. *Public Opin. Q.* **2007**, *71*, 287–311. [CrossRef]
110. Patrick, D.L.; Burke, L.B.; Gwaltney, C.J.; Leidy, N.K.; Martin, M.L.; Molsen, E.; Ring, L. Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1—Eliciting concepts for a new PRO instrument. *Value Health* **2011**, *14*, 967–977. [CrossRef]
111. Janz, K.F.; Lutuchy, E.M.; Wenhe, P.; Levy, S.M. Measuring activity in children and adolescents using self-report. *Med. Sci. Sports Exerc.* **2008**, *40*, 767–772. [CrossRef]
112. Kremers, S.P.J.; Visscher, T.L.S.; Seidell, J.C.; van Mechelen, W.; Brug, J. Cognitive determinants of energy balance-related behaviours. *Sports Med.* **2005**, *35*, 923–933. [CrossRef]
113. Sallis, J.F.; Saelens, B.E. Assessment of physical activity by self-report: Status, limitations, and future directions. *Res. Q. Exerc. Sport* **2000**, *71*, 1–14. [CrossRef] [PubMed]
114. Saris, W.H.M. The assessment and evaluation of daily physical activity in children. A review. *Acta Paediatr.* **1985**, *74*, 37–48. [CrossRef] [PubMed]
115. Caspersen, C.J.; Nixon, P.A.; Durant, R.H. 13 physical activity epidemiology applied to children and adolescents. *Exerc. Sport Sci. Rev.* **1998**, *26*, 341–403. [CrossRef]
116. Sallis, J.F. Self-report measures of children's physical activity. *J. Sch. Health* **1991**, *61*, 215–219. [CrossRef] [PubMed]
117. Oliver, M.; Schofield, G.M.; Kolt, G. Physical activity in preschoolers. *Sports Med.* **2007**, *37*, 1045–1070. [CrossRef] [PubMed]