

Diseño y validación de una escala para medir en profesores de Educación Física el uso responsable de las pruebas de condición física (FITPET)

Design and validation of a scale to assess the responsible use of fitness testing in physical education teachers (FITPET)

David Blanco-Luengo¹, Alberto Nuviala¹, Rocío Izquierdo-Gómez^{2,3}, Alberto Grao-Cruces^{2,3}

1 Departamento de Deporte e Informática, Universidad Pablo de Olavide, España.

2 Grupo de investigación GALENO, Departamento de Didáctica de la Educación Física, Plástica y Musical, Facultad de Ciencias de la Educación, Universidad de Cádiz. España.

3 Instituto de Investigación e Innovación Biomédica de Cádiz (INIICA), Unidad de Investigación. España.

CORRESPONDENCIA:

David Blanco Luengo
dblalue@upo.es

CÓMO CITAR EL ARTÍCULO:

Blanco-Luengo, D., Nuviala, A., Izquierdo-Gómez, R., & Grao-Cruces, A. (2020). Diseño y validación de una escala para medir en profesores de Educación Física el uso responsable de las pruebas de condición física (FITPET). *Cultura, Ciencia y Deporte*, 15(46), 551-560.

Recepción: junio 2020 • Aceptación: octubre 2020

Resumen

Objetivo: el presente estudio tuvo como objetivo diseñar y validar una escala para evaluar en profesores de Educación Física el uso responsable de las pruebas de condición física relacionadas con la salud. **Método:** participaron un total de 374 docentes de secundaria (266 hombres y 108 mujeres) de centros educativos pertenecientes a las ocho provincias andaluzas, con edades comprendidas entre 24 y 67 años (41.92 ± 9.24 años). **Resultados:** en primer lugar, se diseñó un instrumento de 26 ítems, que obtuvo valores de correlación ítem-total $\geq .35$, exceptuando 5 ítems. Posteriormente, se evaluó la estructura interna de la escala, resultando esta pertinente. El análisis estadístico de los ítems y el posterior análisis factorial exploratorio dio como resultado cinco factores (orientación del resultado, uso de estilos participativos, el enfoque educativo, rigor metodológico y refuerzo positivo), con correlaciones positivas y significativas, constatadas mediante análisis factorial confirmatorio, pruebas de invarianza factorial, así como de fiabilidad y validez convergente y discriminante. Los resultados del análisis de propiedades psicométricas de la escala fueron adecuados y permitieron ofrecer un instrumento ad hoc de 15 ítems, con moderados valores de validez y fiabilidad, acreditando su utilidad para valorar en profesores de Educación Física el uso responsable de las pruebas de condición física en sus clases.

Palabras clave: condición física, enseñanza, evaluación formativa, adolescentes, capacidad cardiorrespiratoria.

Abstract

Objective: The aim of this study was to design and validate a scale that would allow the evaluation in Physical Education teachers the responsible use of physical fitness tests related to health. **Method:** The sample was composed of 374 secondary school teachers (266 men and 108 women), from educational centers belonging to the eight Andalusian provinces, aged between 24 and 67 years (41.92 ± 9.24). **Results:** Firstly, a 26-item instrument was designed, which obtained item-total correlation values of $\geq .35$, except for 5 items. Then, the internal structure of the scale was evaluated and proved to be relevant. The statistical analysis of the items and the subsequent exploratory factor analysis resulted in five factors (orientation of the result, use of participatory styles, the educational approach, methodological rigor, and positive reinforcement), with positive and significant correlations, confirmed by confirmatory factor analysis, factor invariance tests, as well as reliability and convergent and discriminant validity and reliability. The results of the analysis of psychometric properties of the scale were moderate and allowed us to offer a valid and reliable 15-item ad hoc instrument, with adequate values of validity and reliability, proving its usefulness the evaluation in Physical Education teachers the responsible use of physical fitness tests in their classes.

Key words: physical fitness, teaching, formative evaluation, teenagers, cardiorespiratory capacity.

Introducción

La condición física (CF) es la capacidad de realizar actividad física, y tiene en cuenta toda la gama de cualidades fisiológicas y psicológicas (Ortega et al., 2008). La evaluación de la CF es fundamental, ya que se puede considerar una medida integrada de gran parte de las funciones del cuerpo y al mismo tiempo un marcador importante de salud, especialmente en edades tempranas, por los cambios que se producen a estas edades y las consecuencias que tienen en la edad adulta (Ortega et al., 2008). En este contexto, el entorno educativo es un espacio favorable para el desarrollo de hábitos saludables y detectar posibles problemas de salud. Por tanto, la evaluación de la CF en el ámbito escolar, especialmente durante las clases de Educación Física (EF), es un contenido tradicional en los currículos educativos de países europeos y norteamericanos (Forest et al., 2018; Hastie, 2017; Keating et al., 2013). En España, el Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato, contempla la promoción de la salud y el desarrollo de la competencia motriz como las finalidades educativas de la EF en educación secundaria. En esa línea, la normativa que desarrolla la Ley Orgánica 8/2013 para la mejora de la calidad educativa (LOMCE) en las comunidades autónomas aluden de forma explícita al desarrollo y evaluación de la CF en sus elementos curriculares, si bien es cierto que el tratamiento educativo de este contenido supone todavía uno de los retos de la EF en pro de la salud del alumnado (Pastor et al., 2016).

Según Bouchard et al. (2012), los componentes de la CF relacionada con la salud son la capacidad muscular, motora, cardiorrespiratoria y la composición corporal. La capacidad cardiorrespiratoria y muscular, junto con la composición corporal, son componentes prioritarios de la CF relacionada con la salud, debido a su validez predictiva de la misma (Ruiz et al., 2009). Un correcto desarrollo de estas capacidades se asocia con un perfil metabólico-lipídico saludable, una mayor sensibilidad a la insulina y tolerancia a la glucosa, una reducción de la presión arterial, una mayor densidad ósea, una composición corporal saludable y un mayor rendimiento cognitivo y académico en adolescentes (Ruiz-Ariza et al., 2017; Ruiz et al., 2016; Smith et al., 2014). De ahí la importancia de detectar niveles poco saludables de capacidad cardiorrespiratoria, capacidad muscular y composición corporal en jóvenes en edad escolar (Ruiz et al., 2016). Para ello se cuenta con una batería de pruebas de CF relacionada con la salud, la ALPHA-*fitness test battery* (Ruiz et al., 2011),

financiada con fondos europeos que ha demostrado su fiabilidad, viabilidad y seguridad para ser aplicada por el docente de EF en los centros educativos, tal y como recomienda la Comisión Europea (España-Romero et al., 2008). Además, recientemente están proliferando estudios de alto rigor científico con puntos de corte relacionados con la salud en las pruebas que componen esta batería para población escolar (Castro-Piñero et al., 2017, 2019; Ruiz et al., 2016).

Un aspecto que ha aumentado la preocupación sobre los niveles de CF basados en la salud de los estudiantes fue el aumento progresivo del sobrepeso y la obesidad en los jóvenes durante las últimas dos décadas (Ogden et al., 2016). Sin embargo, la evaluación de la CF en el entorno educativo es un elemento controvertido del currículo de EF, preocupando su posible efecto desmotivador en los estudiantes con bajo rendimiento físico (Grao-Cruces et al., 2020; Jaakkola et al., 2013; Wiersma & Sherman, 2008). Dentro de este colectivo se encuentran presumiblemente los estudiantes con sobrepeso y obesidad, cuyos niveles promedios de CF suele ser inferior al de sus pares sin exceso de peso (Martínez-López, de la Torre-Cruz, Suárez-Manzano & Ruiz -Ariza, 2017; Palomäki, Heikinaro-Johansson & Huotari, 2015). Existe preocupación por las consecuencias negativas a nivel psicológico que las pruebas de CF podrían tener entre el alumnado con exceso de peso si el contenido no se trata adecuadamente (Lodewyk & Sullivan, 2016; Martínez-López, Grao-Cruces, Moral-García & De la Torre Cruz, 2013).

En este sentido, el debate sobre los diferentes enfoques para aplicar estas pruebas en el entorno educativo, que no es nuevo, sigue vigente actualmente (Cale et al., 2007; Corbin et al., 1995; Rowland, 1995). Algunos investigadores sugirieron suspender las pruebas de CF de las programaciones docentes de EF (Rowland, 1995), alegando que el uso de estas pruebas se vinculaba al rendimiento físico, cuando la tendencia era una EF orientada a la salud (Corbin et al., 1995; Rowland, 1995). Otros estudios propusieron centrarse en los aspectos educativos que podrían aportar las pruebas de CF y vincularlas a la enseñanza de las mismas, la interpretación de sus resultados, su repercusión sobre la salud y los medios para la mejora de las capacidades que evalúan (Cale & Harris, 2002; Corbin & Pangrazi, 1993; Ratliffe & Ratliffe, 1994) dentro de una EF integral (Rowland, 1995). En línea con esta última perspectiva, diversos autores han elaborado una serie de recomendaciones metodológicas sobre cómo usar las pruebas de CF en las clases de EF de forma responsable (Silverman et al., 2008; Wiersma & Sherman, 2008) o pautas prácticas para eliminar el énfasis de la prueba y centrarse en el proceso y la promoción de la CF relacio-

nada con la salud (Meredith & Welk, 2013; Presidential Youth Fitness Program [PYFP], 2013).

Sería de interés el desarrollo de un instrumento capaz de medir este uso responsable de las pruebas de CF por el profesorado de EF con el objetivo de favorecer la reflexión entre el profesorado y unificar criterios en el tratamiento educativo de este contenido. Hasta donde conocen los autores, no existen instrumentos validados que evalúen el uso de estas pruebas por parte del profesorado. Por ello el presente estudio tiene como objetivo diseñar y validar un instrumento para medir en el profesorado de EF, de manera válida y fiable, el uso responsable de las pruebas de CF relacionada con la salud durante sus clases.

Método

Participantes

La muestra estuvo formada por 374 docentes de EF (266 hombres y 108 mujeres) de entre 24 y 67 años (41.92 ± 9.24 años) pertenecientes a centros educativos públicos (327), concertados (38) y privados (9) (ESO) de la Comunidad Autónoma de Andalucía y experiencia docente con un rango de 1 a 44 años (15.44 ± 9.60 años). La muestra seleccionada fue representativa de los docentes de EF de ESO en centros educativos andaluces, con un error $< .03$ a un nivel de confianza del 95%.

Diseño e Instrumento

El diseño de este instrumento se articuló en cuatro fases consecutivas: a) búsqueda bibliográfica: en la primera fase se revisaron los instrumentos previamente publicados, utilizando para ello las bases de datos PubMed, Psycodoc, ERIC y SportDiscus, e introduciendo la ruta de búsqueda: fitness testing en el apartado de Título/Resumen; b) diseño de instrumento inicial: la versión preliminar de la escala, elaborada tras un trabajo cualitativo que sigue las normas establecidas por Carretero-Dios y Pérez (2005) sobre la construcción de instrumentos de investigación, estuvo conformada por 26 ítems. Para la selección de los ítems se recopiló un listado de recomendaciones para el uso de las pruebas de CF en la escuela (Silverman et al., 2008; Wiersma & Sherman, 2008). Con objeto de asegurar tanto la validez de contenido como la aplicabilidad del instrumento esta primera versión fue sometida a un doble proceso de depuración (juicio de expertos y estudio piloto): c) juicio de expertos: en la tercera fase, seis participantes analizaron la ade-

cuación de los ítems con la dimensión que evaluaban mediante una escala Likert de cinco puntos, utilizándose el acuerdo inter-jueces para eliminar o depurar los ítems problemáticos; la selección de los expertos se fundamentó en dos criterios: primero que fueran especialista en la materia de EF, valorándose por ello su producción bibliográfica y haber participado con anterioridad en el diseño de instrumentos. Así, los principales cambios derivados de la validación de contenido fueron exclusivamente gramaticales. d) Estudio piloto: participaron 20 sujetos con el objeto de depurar los ítems que más problemas de comprensión generaban o que presentaban erratas en su formulación. Los 26 ítems iniciales superaron dicho proceso y fueron utilizados (Tabla 1).

El estudio siguió un diseño instrumental (Montero & León, 2007). Se elaboró una escala para evaluar el uso responsable de las pruebas de condición física por el profesorado de EF (FITPET). Este instrumento presenta finalmente variables sociodemográficas junto a 15 ítems, cada ítem se valoró mediante una escala Likert de cinco puntos en la que 1 era totalmente en desacuerdo y 5 totalmente de acuerdo (Anexo 1).

Procedimiento

La participación de la muestra fue voluntaria, consentida y anónima; cada docente, en el momento de pasar el cuestionario no impartía clase. En primer lugar, se informó a los responsables de los centros educativos que participaron en el estudio, tras su aprobación se realizó la investigación. Se ha tenido en cuenta la normativa legal vigente de España que regula la protección de datos de carácter personal (Ley Orgánica 3/2018), así como los principios fundamentales establecidos en la Declaración de Helsinki (revisión de 2013, Brasil). La escala fue elaborada y facilitada a los docentes por medios físicos y electrónicos en presencia de un investigador.

Análisis de datos

En primer lugar, se desarrolló un análisis estadístico descriptivo de los ítems. En segundo lugar, se procedió al Análisis Factorial Exploratorio (AFE) con el objetivo de comprobar la agrupación de los ítems empíricamente. Se confirmó la existencia de correlaciones significativas para la adecuación de la matriz mediante el test Kaiser-Meyer-Olkin (KMO) y la Prueba de Esfericidad de Bartlett, comprobándose la adecuación de utilizar AFE.

Seguidamente se realizó un análisis factorial confirmatorio (AFC). El método utilizado fue el de máxima

Tabla 1. Media (M), desviación típica (DT), Asimetría, Curtosis, correlación ítem-total (R IT-c) de los 26 ítems iniciales.

Ítem	M	DT	Asimetría	Curtosis	R IT-c
1	4.09	.90	-.91	.63	.46
2	4.59	.73	-1.57	1.89	.36
3	4.09	.98	-1.04	0.48	.46
4	3.59	1.06	-.56	-.25	.58
5	3.66	1.01	-.59	-.12	.22
6	4.39	.71	-1.36	1.96	.51
7	3.22	1.17	-.11	-.92	.29
8	3.84	.82	-.56	.13	.38
9	3.86	.88	-.69	.41	.51
10	3.53	.87	-.42	.23	.53
11	3.50	.97	-.40	-.36	.28
12	3.93	.75	-.84	1.62	.40
13	3.77	.79	-.56	.60	.50
14	3.80	.80	-.65	.82	.55
15	4.22	.65	-.60	1.29	.49
16	3.73	.84	-.64	.46	.22
17	3.61	1.00	-.59	-.20	.36
18	3.74	.98	-.70	-.02	.27
19	3.92	.94	-.88	.69	.44
20	4.28	.72	-.95	1.07	.53
21	3.23	1.20	-.39	-.83	.42
22	4.52	.60	-1.29	1.86	.42
23	3.94	.90	-.84	.54	.52
24	4.20	.74	-1.24	1.95	.49
25	4.01	.93	-.96	.65	.44
26	4.58	.62	-1.88	1.91	.40

*Nota: Media (M), desviación típica (DT), correlación ítem-total (R IT-c).

verosimilitud. Para evaluar la bondad del ajuste, se revisaron los siguientes indicadores: chi cuadrado de Pearson (χ^2); grados de libertad (gl); índice de ajuste comparativo (CFI); Error de Aproximación Cuadrático Medio (RMSEA); índice de error de cuadrático medio (RMR) (Hair, Black, Babin, Anderson y Tatham, 2006; Marôco, 2010). Además, con el objeto de seguir las indicaciones de Byrne (2001) se adjunta el criterio de información de Akaike (AIC) y el índice de validación cruzada esperada (ECVI). A continuación, se calculó la invarianza factorial con el objeto de comprobar la estabilidad del modelo en diferentes poblaciones, de acuerdo a la variable género.

Se realizaron pruebas de validez convergente mediante el cálculo de correlaciones entre factores y cálculo de la fiabilidad, mediante alfa de Cronbach y fiabilidad compuesta (FC). Para concluir se comprobó la validez discriminante mediante el cálculo de correlaciones entre factores y comparación con la raíz cuadrada de la varianza media extraída (AVE). El tratamiento estadístico se llevó a cabo con los paquetes estadísticos IBM SPSS Statistics 21,0 y IBM SPSS AMOS 21,0 (IBM Software Group, Armonk, NY, US). El nivel de confianza establecido fue del 95% ($p < .05$).

Tabla 2. Estructura factorial rotada, comunalidades, autovalores, alfa de Cronbach y porcentaje de varianza explicada por cada factor.

Ítem	Componente					Extracción
	1	2	3	4	5	
20	.75					.64
24	.72					.59
23	.70					.58
19	.64					.47
6	.40					.39
13		.76				.62
14		.73				.62
12		.65				.47
10		.62				.54
9		.62				.49
8		.59				.36
1			.77			.69
2			.67			.61
3			.65			.62
25				.72		.58
17				.72		.56
21				.52		.49
4				.52		.65
26					.75	.63
22					.73	.65
15					.55	.48
% Varianza explicada	14.59	14.48	9.01	8.94	8.55	55.64
Autovalor	6.11	1.89	1.42	1.18	1.05	

Resultados

Análisis estadístico de los ítems

La tabla 1 muestra los estadísticos descriptivos de los ítems. Los valores de correlación ítem-total fueron $\geq .35$, exceptuando los ítems 5, 7, 11, 16 y 18, eliminados por incumplimiento de correlación ítem-total. Se puede observar que los índices de asimetría y curtosis estuvieron por debajo o son iguales al valor 1.96, lo que indica semejanza con la curva normal (Bollen & Long, 1993). Estos resultados permiten la utilización de técnicas factoriales que se realizarán a continuación.

Análisis de la estructura interna

Para conocer la estructura factorial de la escala, se realizó un AFE sobre los 21 ítems resultantes del análisis estadístico. Antes de realizar el análisis se calculó la medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) y el test de esfericidad de Bartlett. El índice KMO mostró un valor de .88 y el test de Bartlett resultó estadísticamente significativo

Tabla 3. Indicadores de ajuste del análisis factorial confirmatorio, comparación entre modelos usando el modelo 1 como correcto,

Modelo	χ^2	GI	χ^2/gl	CFI	RMSEA	RMR	AIC	ECVI	Dif. GL	Dif. χ^2	P
Modelo 0a	121.34	80	1.52	.95	.04	.03	201.34	.76			
Modelo 0b corregido	123.62	80	1.56	.96	.05	.04	203.62	.61			
Modelo 1	253.79	160	1.59	.93	.04	.05	413.79	1.28			
Modelo 2	275.92	170	1.62	.92	.04	.07	415.92	1.28	10	22.13	.14
Modelo 3	312.52	185	1.70	.91	.05	.09	423.52	1.31	25	59.73	.00
Modelo 4	338.55	200	1.69	.90	.05	.09	418.54	1.29	40	84.75	.00

($\chi^{210} = 2328,045$; $p < .001$), lo que llevó a concluir que la aplicación del análisis factorial resultaba pertinente. Los factores conjuntamente explican un 55.64% de la varianza (tabla 2).

Análisis factorial confirmatorio e invarianza factorial

El análisis estadístico de los ítems y el posterior AFE dio como resultado cinco factores: i) orientación del resultado, ii) uso de estilos participativos, iii) enfoque educativo, iv) rigor metodológico y v) refuerzo positivo. El primero de los factores recoge información acerca de la orientación y uso de los resultados de los profesores de EF en las pruebas de CF. En los ítems de esta dimensión aparecen orientaciones e información a facilitar al alumnado, o utilización de los resultados para evaluar el estado físico del sujeto, relacionado con la salud. El segundo de los factores hace alusión a la utilización de estilos de enseñanza participativos durante la prueba y su relación con la reducción de miedo escénico, percepción sobre el propio estado físico e intenciones futuras de realizar actividad física. El tercero de los factores, enfoque educativo, hace alusión a la promoción del disfrute y sus beneficios en la obtención de experiencias positivas en las pruebas y en la adquisición de estilos de vida activos, así como la orientación de la prueba de CF, hacia la promoción de práctica regular de actividad física. El factor rigor metodológico incluye una serie de ítems relativos a la fiabilidad, estandarización de criterios e indicadores de evaluación del proceso en las pruebas utilizadas. Finalmente, el factor refuerzo positivo aborda con sus ítems la motivación, feedback y refuerzo utilizado por el profesorado independientemente del resultado de la prueba. Estos factores explican el 55.64% de la varianza, alrededor de los cuales se agrupaban los ítems.

Los parámetros fueron estimados mediante el método de máxima verosimilitud, con la búsqueda de valores de CFI $> .90$, menores o iguales a $.06$ en RMSEA y RMR, menores o iguales a 2.00 en χ^2/gl , lo más lejanos posibles a 0 en AIC y el menor valor de ECVI. En la tabla 3 se recoge la información proporcionada por

los índices de ajuste utilizados para la escala resultante del AFE compuesta por 5 factores y 21 ítems (modelo 0a). Tras un ajuste, fueron eliminados los ítems 6, 8, 9, 12, 17 y 19, dando lugar al modelo final configurado con 5 factores y 15 ítems, obteniendo índices satisfactorios (modelos 0b corregido).

Posteriormente se analizó la invarianza de esta estructura factorial a través del análisis multigrupo de acuerdo a la variable género. La tabla 3 muestra la inexistencia de diferencias significativas entre un modelo sin invarianza (modelo 1) y el resto de modelos con invarianza en χ^2 . Al examinar los valores de CFI se observa que son semejantes con mínimas diferencia, por lo que se puede sugerir que la estructura del modelo es invariante (tabla 3).

Modelo 0a (hombres), resultante del Análisis Factorial exploratorio; Modelo 0b modelo corregido (mujeres); Modelo 1, que no tiene restricciones de ningún tipo; Modelo 2, tiene restricciones en el peso de medida; Modelo 3, tiene restringidos los pesos de medida y covarianzas; Modelo 4, tiene restricciones en los pesos de medida, covarianzas y residuos de medida. Chi cuadrado (prueba χ^2 de Pearson), grado de libertad (GI), índice de ajuste comparativo (CFI), Error de Aproximación Cuadrático Medio (RMSEA), índice de error de cuadrático medio (RMR), criterio de información de Akaike (AIC), índice esperado de validación cruzada (ECVI).

Validez y fiabilidad

Se constató la validez convergente a través del cálculo de coeficientes de correlación de Pearson entre los factores de la escala. Los resultados ponen de manifiesto la existencia de correlaciones positivas y significativas entre los factores que la conforman. La validez discriminante, medida con la raíz cuadrada de AVE, mostró valores superiores a las correlaciones entre factores que conforman el cuestionario; es decir, que realmente es una validez discriminante (tabla 4). La fiabilidad de los factores viene avalada por los valores de alfa comprendidos entre $.64$ y $.77$ y valores de FC superiores todos a $.70$.

Tabla 4. Correlaciones entre los factores de FITPET, alfa de Cronbach, Fiabilidad Compuesta (FC) y Raíz Cuadrada de la Varianza media Extraída (AVE) (en la diagonal).

Factores	Ítems	Fact1	Fact2	Fact3	Fact4	Fact5	α	FC
Fact1	20							
	23	(.82)	.36**	.32**	.53**	.45**	.76	.86
	24							
Fact2	10							
	13		(.83)	.41**	.36**	.36**	.77	.87
	14							
Fact3	1							
	2			(.77)	.35**	.41**	.64	.81
	3							
Fact4	4							
	21				(.77)	.36**	.66	.81
	25							
Fact5	15							
	22					(.76)	.64	.81
	26							

** Correlación significativa al nivel $p < .05$ (bilateral).

Discusión

El presente estudio tuvo como objetivo diseñar y validar un instrumento para medir de manera válida y fiable el uso responsable de las pruebas de CF relacionadas con la salud durante las clases de EF. Se presenta la escala FITPET que permite conocer la orientación del resultado, uso de estilos participativos, el enfoque educativo, rigor metodológico y refuerzo positivo de los docentes de EF en materia de utilización de pruebas de evaluación de la CF durante las clases de EF con resultados moderados de validez y fiabilidad. Las emanaciones del estudio presentan una escala compuesta de 15 ítems que maximizan la varianza del test, seleccionando aquellos con un elevado poder de discriminación y con puntuaciones medias de respuesta situadas en torno al punto medio de la escala (Bollen & Long, 1993; Carretero-Dios & Pérez, 2005; Nunnally & Bernstein, 1995), dando lugar a un instrumento original e inédito.

En primer lugar, se desarrolló un análisis estadístico descriptivo de los 26 ítems de la escala inicial, con el objetivo de conocer la correlación ítem-total (Cohen & Manion, 1990) y la fiabilidad de la escala a través del alfa de Cronbach. En segundo lugar, se procedió al AFE con el objetivo de comprobar la agrupación de los ítems empíricamente. De esta manera se pudo explorar la estructura interna de la escala su dimensionalidad o forma de agruparse los ítems (Elosua, 2003). Dicho análisis proporcionó los agrupamientos de las variables que componen el cuestionario en función de criterios matemáticos basados en la correspondencia entre estos para que posteriormente sean interpretados. El AFE agrupa solo correlaciones similares, pero esta agrupación puede ser debida a más elementos

que los propiamente conceptuales. El procedimiento de rotación utilizado fue Varimax, a pesar de que se aconseja para casos en que los factores no están relacionados. Se optó por este debido al interés teórico de separar, en la medida de lo posible, los factores resultantes, a pesar de constatar la relación de los factores (Carretero-Dios & Pérez, 2007).

Previo a una correcta aplicación del AFE es necesario contrastar que los ítems deben encontrarse relacionados entre sí; es decir, la matriz de correlaciones debe ser tal que puedan localizarse agrupamientos relevantes entre variables. Por ello es necesario realizar antes de la aplicación del análisis el cálculo de unos estimadores que aseguren que la matriz de correlaciones es la apropiada (Cortina, 1993), siendo las pruebas de elección la de esfericidad de Bartlett y el índice de Kaiser-Meyer-Olkin (KMO). El resultado de esta prueba dio la pertinencia del proceso al presentar un valor alto.

Respecto al AFC para constatar la estructura factorial de la escala, los parámetros fueron estimados mediante el método de máxima verosimilitud (Thompson, 2004). Para evaluar la adecuación del modelo sometido a prueba se optó por la valoración conjunta de un grupo de índices, los cuales fueron seleccionados de algunos de los índices de ajuste más utilizados, considerándose aceptables valores en el caso del CFI, por encima de .90. En el caso del RMSEA y RMR, el modelo presentaría un ajuste aceptable si el valor fuera $< .07$ (Steiger, 2007); valores $\leq .06$ indicarían un buen ajuste (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Por lo que se refiere a los valores del cociente entre χ^2 y los gl , un modelo considerado perfecto su valor sería de 1.00 y ratios por debajo de 2.00 se considerarán de un muy buen ajuste del modelo, mientras que valores

por debajo de 5.00 son considerados como aceptables (Hu & Bentler, 1999; MacCallum, Widaman, Preacher, & Hong, 2001; Yuan, 2005). Finalmente, debido a la conveniencia de comparar el ajuste del modelo, se incorporaron dos índices especialmente desarrollados para ello. El índice AIC, índice comparativo entre modelos, debiendo elegir el modelo que presente un menor valor AIC (Schermelleh-Engel et al., 2003); valores más cercanos a cero indican un mejor ajuste. El índice ECVI (*Expected Cross Validation Index*) mide la discrepancia entre la matriz de covarianzas implicada en la muestra analizada y la matriz de covarianzas esperada para otra muestra del mismo tamaño. Cuando se están comparando modelos, el menor valor de ECVI indica el modelo con mejor ajuste (Schermelleh-Engel et al., 2003). Los resultados de los diferentes índices de ajuste del modelo original pueden ser considerados como correctos. Sin embargo, tras las modificaciones sugeridas por el programa, el modelo corregido presentó mejores índices de ajuste. Además, el instrumento obtuvo valores de fiabilidad entre .64 y .77 y de FC superiores a .70 para todos los factores.

Posteriormente se analizó la invarianza de la estructura factorial a través del análisis multigrupo (Abalo, Lévy, Rial, & Varela, 2006). Para ello se dividió el grupo en dos subgrupos de profesores en función del género. Se trataba de comprobar que no hubiera diferencias significativas entre un modelo sin invarianza y diferentes modelos con invarianza en algunos parámetros. No se encontraron diferencias significativas en chi-cuadrado entre el modelo sin restricciones (Modelo 1) y el resto de modelos. Además, teniendo presente que el coeficiente chi-cuadrado es sensible al tamaño de la muestra, se empleó también el criterio establecido por Cheung y Rensvold (2002) respecto al ΔCFI .

La validez convergente se determinó por las correlaciones entre los 5 factores que conforman la escala FITPEC a través del coeficiente de Pearson. Las correlaciones entre ellos fueron calificadas desde cuestionables (George & Mallery, 2003) o de fiabilidad moderada (DeVellis (1991) a aceptables (George & Mallery, 2003). Los resultados de las correlaciones requieren una segunda prueba de validez convergente del instrumento, determinada con el índice de fiabilidad (FC). Los valores aceptables son $>.70$ (Fornell & Larcker, 1981; Bagozzi & Yi, 1988). Ambas pruebas ponen de manifiesto la existencia de este tipo de validez.

La validez discriminante de la escala viene expresada por el contraste entre los diferentes factores que la componen. Si son realmente distintos los conceptos que la conforman y al mismo tiempo están relacionados es entonces cuando se puede hablar de este tipo de validez (Lehmann, Gupta, & Steckel, 1999). Para ga-

rantizarla se comparó la raíz cuadrada de la AVE con la correlación entre constructos que conforman la escala (Fornell & Larcker, 1981). Para que exista validez discriminante entre constructos es necesario que la raíz cuadrada de la AVE sea superior a la correlación entre los mismos. Considerando los resultados de las correlaciones y viendo los valores de AVE se puede afirmar que existe una validez discriminante.

El instrumento que se presenta es inédito, diseñado y validado en idioma español, dirigido a docentes de EF de secundaria y de rápida aplicación. Son pocas las referencias que encontramos sobre la validación de cuestionarios dedicados a medir el uso responsable de las pruebas de CF por el profesorado de EF. No obstante, se pueden destacar en la literatura científica publicaciones que proporcionan evaluaciones, promoción de actividades y sistema de retroalimentación para estudiantes, profesores y padres con el fin de fomentar la CF relacionada con la salud durante toda la vida (Meredith & Welk, 2013) y otras pautas prácticas de evaluación de la CF para profesores de EF, a través de listas de comprobación que recogen componentes como ayudar a los estudiantes a establecer metas relacionadas con la salud, brindar múltiples formas de evaluación (formal, por pares, autoevaluación), revisión de protocolos por parte de los profesores o el refuerzo en cada evaluación del componente de salud que mide y actividades físicas que pueden influir en él (Presidential Youth Fitness Program [PYFP], 2013). Estas buenas prácticas están contempladas y valoradas positivamente en algunos de los factores que componen el instrumento diseñado y validado por el presente estudio.

Sin embargo, el estudio presenta algunas limitaciones, como la moderada fiabilidad de algunos factores que conforman la escala, la imposibilidad de poder contrastarlo con estudios previos y la inexistencia de estabilidad temporal en relación al contenido impartido. La cuestionable fiabilidad de alguno de los factores puede tener relación con la existencia de algún ítem alejado de la representatividad del constructo o la interpretación no correcta de las preguntas, dificultando la comprensión del entrevistado que utilizó el medio electrónico, todo ello minimizado por el carácter anónimo del mismo y la validez obtenida. La estabilidad temporal de instrumento podría subsanarse en futuras investigaciones, realizando test-retest con un tiempo de diferencia mínimo de una semana. En cualquier caso, con independencia del sexo, temporalidad, experiencia del docente, y aun cuando el cuestionario cuenta con instrucciones específicas sobre cómo deben formularse las respuestas, existe la posibilidad de que una presencia más activa del encuestador pueda ayudar a mejorar la fiabilidad del instrumento.

Conclusiones

En conclusión, los resultados posibilitan el ofrecer una escala inédita que permite conocer la orientación del resultado, uso de estilos participativos, enfoque educativo, rigor metodológico y refuerzo positivo de los profesores de EF en materia de evaluación

de las pruebas de CF durante las clases de Educación Física. Tras los análisis factoriales se presenta un instrumento ad hoc de 15 ítems, con moderados valores de validez y fiabilidad, que acredita su utilidad para el uso responsable del profesorado en centros educativos, permitiendo la comparación de posibles resultados.

Anexo 1. Escala para evaluar el uso responsable de las pruebas de condición física por el profesorado de EF (FITPET).

Queremos conocer su OPINIÓN sobre el uso responsable de las pruebas de condición física durante las clases de Educación Física. No hay respuestas buenas ni respuestas malas, se trata únicamente de conocer su OPINIÓN más sincera. El trabajo cumple rigurosamente con los principios éticos de aplicación en nuestro país. Los datos obtenidos tendrán carácter totalmente anónimo y confidencial.

Descripción	1	2	3	4	5
1 Un enfoque de la Educación Física que promueve el disfrute prepara al alumnado para una experiencia positiva en las pruebas de condición física.	1	2	3	4	5
2 Un enfoque de la Educación Física que promueve el disfrute prepara al alumnado para la adquisición de un estilo de vida activo.	1	2	3	4	5
3 Una finalidad de las pruebas de condición física debe ser la promoción de la práctica regular de actividad física.	1	2	3	4	5
4 La utilización de estilos de enseñanza que fomentan la participación durante las pruebas de condición física disminuye el miedo escénico, evitando una experiencia negativa.	1	2	3	4	5
5 La percepción del alumnado sobre su estado físico mejora con la utilización de estilos de enseñanza que fomentan la participación durante las pruebas de condición física.	1	2	3	4	5
6 La utilización de estilos de enseñanza que fomentan la participación durante las pruebas de condición física, condiciona positivamente las intenciones futuras de realizar actividad física.	1	2	3	4	5
7 El alumnado que ponen el énfasis en la superación personal, se desenvuelven mejor cuando se refuerza la mejora de las habilidades.	1	2	3	4	5
8 El profesorado debe usar criterios estandarizados para proporcionar retroalimentación al alumnado.	1	2	3	4	5
9 Los resultados de las pruebas de condición física se deben usar para ayudar al alumnado a planificar el mantenimiento o mejora de la condición física relacionado con la salud.	1	2	3	4	5
10 El profesorado debe valerse de los resultados de las pruebas de condición física como uno de los indicadores para evaluar el proceso de E-A en relación al contenido condición física.	1	2	3	4	5
11 El profesorado debe favorecer la motivación hacia la condición física para fomentar hábitos de vida activos.	1	2	3	4	5
12 A través de las pruebas de condición física, el profesorado puede evaluar el estado físico relacionado con la salud.	1	2	3	4	5
13 Con las pruebas de condición física, el profesorado puede ayudar al alumnado a comprender los componentes de la condición física y sus interrelaciones.	1	2	3	4	5
14 El profesorado debe utilizar pruebas de condición física cuya validez y fiabilidad se haya contrastado previamente.	1	2	3	4	5
15 El profesorado debe proporcionar feedback sobre el esfuerzo y/o mejora de la condición física, independientemente del resultado de la prueba.	1	2	3	4	5

*Nota: Totalmente en desacuerdo (1), en desacuerdo (2), ni de acuerdo ni en desacuerdo (3), de acuerdo (4), totalmente de acuerdo (5). Versión final exceptuando los ítems 5, 7, 11, 16 y 18, eliminados por incumplimiento de correlación ítem-total y los ítems 6, 8, 9, 12, 17 y 19, eliminados en el ajuste final del AFE de la versión original de 26 ítems.

BIBLIOGRAFÍA

- Abalo, J., Lévy, J. P., Rial, A., & Varela, J. (2006). Invarianza factorial con muestras múltiples. En J. P. Lévy y J. Varela (Eds.), *Modelización con Estructuras de Covarianzas en Ciencias Sociales* (pp. 259-278). Madrid: Netbiblo.
- Alarcón, T., & Reyno, A. (2009). Estilos de enseñanza en educación física: Estudio transversal. *Habilidad Motriz*, 33, 15-24.
- Bagozzi, R., & Yi, Y. (1988). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 16 (1), 74-94. <https://doi.org/10.1007/bf02723327>
- Barbero, M. (2003). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.
- Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. Sage: Newbury Park, CA.
- Bouchard, C., Blair, S. N., & Haskell, W. (eds). (2012). *Physical Activity and Health*. 2nd. US: Human Kinetics.
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programmin*. Londres: LEA.
- Cale, L., & Harris, J. (2002). National testing for children: Issues, concerns, and alternatives. *British Journal of Teaching Physical Education*, 33(1), 32-34.
- Cale, L., Harris, J., & Chen, M. H. (2007). More than 10 years after "The Horse is Dead": Surely it must be time to dismount?! *Pediatric Exercise Science*, 19, 115-131. <https://doi.org/10.1123/pes.19.2.115>
- Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and the review of instrumental studies: Considerations about test selection in psychological research, *International Journal of Clinical and Health Psychology*, 7, 863-882.
- Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales, *International Journal of Clinical and Health Psychology*, 5(3), 521-551.
- Castro-Piñero, J., Pérez-Bey, A., Cuenca-García, M., Cabanas-Sánchez, V., Gómez-Martínez, S., Veiga, O. L., Marcos, A., Ruiz, J. R., & UP&DOWN Study Group (2019). Muscle Fitness Cut Points for Early Assessment of Cardiovascular Risk in Children and Adolescents. *The Journal of pediatrics*, 206, 134-141.e3. <https://doi.org/10.1016/j.jpeds.2018.10.026>
- Castro-Piñero, J., Perez-Bey, A., Segura-Jiménez, V., Aparicio, V. A., Gómez-Martínez, S., Izquierdo-Gomez, R., Marcos, A., Ruiz, J. R., & UP&DOWN Study Group (2017). Cardiorespiratory Fitness Cutoff Points for Early Detection of Present and Future Cardiovascular Risk in Children: A 2-Year Follow-up Study. *Mayo Clinic proceedings*, 92(12), 1753-1762. <https://doi.org/10.1016/j.mayocp.2017.09.003>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance, *Structural Equation Modeling*, 9(2), 233-255. http://dx.doi.org/10.1207/S15328007SEM0902_5
- Cohen, L., & Manion, L. (1990). *Métodos de investigación educativa*. Madrid: La Muralla.
- Corbin, C. B., & Pangrazi, R. P. (1995). Physical fitness: Questions teachers ask, *Journal of Physical Education, Recreation and Dance*, 64(7), 14-19. <https://doi.org/10.1080/07303084.1993.10606774>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications, *Journal of Applied Psychology*, 78, 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Delgado, M. A., Medina, J., & Viciano, J. (1996). The teaching styles in the preservice of physical education teachers. *International Seminar. AIESEP*. November. 21-24.
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. California. Sage Publications.
- Elosua, P. (2003). Sobre la validez de los tests, *Psicothema*, 15(2), 315-321.
- Forest, E., Lenzen, B., & Öhman, M. (2018). Teaching traditions in physical education in France, Switzerland and Sweden: A special focus on official curricula for gymnastics and fitness training, *European Educational Research Journal*, 17(1), 71-90, <http://doi.org/10.1177/1474904117708889>.
- Fornell C., & Larcker, D. F. (1981). Evaluating structural model with unobserved variables and measurement errors, *Journal of Marketing Research*, 18 (1), 39-50.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston: Allyn & Bacon.
- González-Peiteado, M., & Pino-Juste, M. (2013b). Percepción del alumnado de Ciencias de la Educación de la Universidad de Santiago de Compostela sobre el uso de los estilos de enseñanza. *Innovación Educativa*, 23, 215-229.
- Grao-Cruces, A.; Racero-García, A.; Sánchez-Oliva, D.; Blanco-Luengo, D.; Nuviala, A.; García-Calvo, T. (2020). Associations between Weight Status and Situational Motivation toward Fitness Testing in Physical Education: The Mediator Role of Physical Fitness. *Int. J. Environ. Res. Public Health*, 17, 4821. <https://doi.org/10.3390/ijerph17134821>
- Hair, J, F, Black, W, C., Babin, B, J., Anderson, R, E., & Tatham, R, L, (2006). *Multivariate data analysis* (6ª Edition), Upper Saddle River, NJ: Pearson-Prentice-Hall.
- Hastie, P. A. (2017). Revisiting the national physical education content standards: What do we really know about our achievement of the physically educated/literate person? *Journal of Teaching in Physical Education*, 36, 3-19. <http://doi.org/10.1123/jtpe.2016-0182>.
- Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, Invalidity and Reliability. In E. F. Borgatta & G. Bohrnstedt (Eds.), *Sociological methodology* (pp.104-129), San Francisco: Jossey Bass.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling*, 6(1), 1-55. <http://dx.doi.org/10.1080/10705519909540118>.
- Jaakkola, T. T., Sääkslahti, A., Yli-Piipari, S., Manninen, M., Watt, A., & Liukkonen, J. (2013). Student motivation associated with fitness testing in the physical education context, *Journal of Teaching in Physical Education*. <https://doi.org/10.1123/jtpe.32.3.270>
- Keating, X. D., & Silverman, S. (2004). Teachers' use of fitness tests in school-based physical education programs, *Measurement in Physical Education and Exercise Science*, 8, 145-165. https://doi.org/10.1207/s15327841mpee0803_2
- Keating, X. D., Subramaniam, P. R., Shanguan, R., & Chen, L. (2013). Physical education program changes from 2006 to 2010, *Journal of Teaching in Physical Education*, 32(2), 205-213. <http://doi.org/10.1123/jtpe.32.2.205>.
- Lehmann, D. R., Gupta, S., & Steckel, J. H. (1999). *Marketing Research*, New York: Addison-Wesley.
- Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales, 6 de diciembre de 2108. BOE n° 294.
- Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa (LOMCE), 10 de diciembre de 2013. BOE n° 295.
- Lodewyk, K. R., & Sullivan, P. (2016). Associations between anxiety, self-efficacy, and outcomes by gender and body size dissatisfaction during fitness in high school physical education, *Physical Education and Sport Pedagogy*, 21(6), 603-615. <http://doi.org/10.1080/17408989.2015.1095869>.
- López Pastor, V. M. & Pérez Brunicardi, D., & Manrique Arribas, J. C., & Monjas Aguado, R. (2016). Los retos de la Educación Física en el Siglo XXI. *RETOS. Nuevas Tendencias en Educación Física, Deporte y Recreación*, (29), 182-187. ISSN: 1579-1726. <https://doi.org/10.47197/retos.v0i29.42552>
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analyses: The role of model error, *Multivariate Behavioral Research*, 36, 611-637. http://dx.doi.org/10.1207/S15327906MBR3604_06.
- Marôco, J. (2010). *Análise de Equações Estruturais. Fundamentos teóricos, Software & Aplicações*. Pero Pinheiro: Rolo y Filhos II, SA.
- Martínez-López, E. J., de la Torre-Cruz, M., Suarez-Manzano, S., & Ruiz-Ariza, A. (2017). Analysis of the effect size of overweight in muscular strength tests among adolescents. Reference values according to sex, age and BMI. *Journal of Strength and Conditioning Research*, 1. <http://doi.org/10.1519/JSC.0000000000001967>.
- Martínez-López, E. J., Grao-Cruces, A., Moral-García, J. E., & de la Torre Cruz, M. J. (2013). Knowledge and attitude. Two key elements in the

- training of physical education teachers to prevent and manage obesity among schoolchildren. *Revista Española de Pedagogía*, 71(256), 525–539.
- Martínez-López, E. J., Zamora-Aguilera, N., Grao-Cruces, A., & de la Torre-Cruz, M. J. (2017). The association between Spanish physical education teachers' self-efficacy expectations and their attitudes toward overweight and obese students. *Journal of Teaching in Physical Education*, 36(2), 220–231. <http://doi.org/10.1123/jtpe.2014-0125>
- Martínez-Vizcaino V. & Sánchez-Lopez M. [Relationship between physical activity and physical fitness in children and adolescents]. *Revista Española de Cardiología*. [https://doi.org/10.1016/s1885-5857\(08\)60084-5](https://doi.org/10.1016/s1885-5857(08)60084-5)
- Meredith, M. D., & Welk, G. J. (2013). *FitnessGram®/ActivityGram test administration manual* (4th ed.). Dallas, TX: Cooper Institute for Aerobics Research.
- Montero, I. & León, O.G. (2007). A guide for naming research studies in psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847–862.
- Nunnally, J.C., & Bernstein, I.J. (1995). *Teoría psicométrica*. Madrid: McGraw-Hill.
- Ogden, C. L., Carroll, M. D., Lawman, H. G., Fryar, C. D., Kruszon-Moran, D., Kit, B. K., & Flegal, K. M. (2016). Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *JAMA*, 315(21), 2292. <http://doi.org/10.1001/jama.2016.6361>.
- Ortega, F. B., Ruiz, J. R., Castillo, M. J., Moreno, L. A., González-Gross, M., Wärnberg, J., Gutiérrez, A., & Grupo AVENA (2005). Bajo nivel de forma física en los adolescentes españoles. Importancia para la salud cardiovascular futura (Estudio AVENA). *Revista española de cardiología*, 58(8), 898–909. <https://doi.org/10.1157/13078126>
- Ortega, F. B., Ruiz, J. R., Castillo, M. J. & Sjostrom, M. (2008). Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes (Lond)*, 32 (1), 1-11. <https://doi.org/10.1038/sj.ijo.0803774>
- Palomäki, S., Heikinaro-Johansson, P., & Huotari, P. (2015). Cardiorespiratory performance and physical activity in normal weight and overweight Finnish adolescents from 2003 to 2010. *Journal of Sports Sciences*, 33(6), 588–596. <https://doi.org/10.1080/02640414.2014.951874>
- Presidential Youth Fitness Program. (2013). *Presidential youth fitness program physical educator resource guide*. Washington, DC: National Foundation on Fitness, Sports and Nutrition.
- Ratliffe, T., & Ratliffe, L. (1994). *Teaching children fitness: Becoming a master teacher*. Champaign, IL: Human Kinetics.
- RD 1105/2014, de 26 de diciembre. Por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. 3 de enero de 2015. BOE N° 3.
- Rowland, T. W. (1995). The horse is dead; let's dismount. *Pediatric Exercise Science*, 7, 117–120. <https://doi.org/10.1123/pes.7.2.117>
- Ruiz-Ariza, A., Grao-Cruces, A., Loureiro, N. E. M. de, & Martínez-López, E. J. (2017). Influence of physical fitness on cognitive and academic performance in adolescents: A systematic review from 2005–2015. *International Review of Sport and Exercise Psychology*, 10(1), 108–133. <https://doi.org/10.1080/1750984x.2016.1184699>
- Ruiz, J. R., Castro-Piñero, J., Artero, E. G., Ortega, F. B., Sjöstrom, M., Suni, J., & Castillo, M. J. (2009). Predictive validity of health-related fitness in youth: a systematic review. *British Journal of Sports Medicine*, 43, 909–923. <http://doi.org/10.1136/bjism.2008.056499>
- Ruiz, J. R., Castro-Piñero, J., España-Romero, V., Artero, E. G., Ortega, F. B., Cuenca, M. M., Jiménez-Pavón, D., Chillón, P., Girela-Rejón, M. J., Mora, J., Gutiérrez, A., Suni, J., Sjöstrom, M., & Castillo, M. J. (2011). Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *British Journal of Sports Medicine*, 45(6), 518–524. <https://doi.org/10.1136/bjism.2010.075341>
- Ruiz, J. R., Cavero-Redondo, I., Ortega, F. B., Welk, G. J., Andersen, L. B., & Martínez-Vizcaino, V. (2016). Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents; what level of fitness should raise a red flag? A systematic review and meta-analysis. *British Journal of Sports Medicine*, 50(23), 1451–1458. <http://doi.org/10.1136/bjsports-2015-095903>
- Ruiz, J., España, V., Castro, J., Artero, E., Ortega, F., Cuenca, M., Jiménez, D., Chillón, P., Girela, M.A J., Mora, J., Gutiérrez, A., Suni, J., Sjöstrom, M., & Castillo, M.J. (2011). Batería ALPHA-Fitness: test de campo para la evaluación de la condición física relacionada con la salud en niños y adolescentes. *Nutrición Hospitalaria*, 26(6), 1210-1214. <https://doi.org/10.3305/nh.2011.26.6.5270>
- Silverman, S., Keating, X. D., & Phillips, S. R. (2008). A lasting impression: A pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science*, 12(3), 146–166. <http://doi.org/10.1080/10913670802216122>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. <https://doi.org/10.1016/j.paid.2006.09.017>
- Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis. Understanding concepts and applications*. Washington, D.C.: American Psychological Association. <http://dx.doi.org/10.1037/10694-000>
- Wiersma, L. D., & Sherman, C. P. (2008). The responsible use of youth fitness testing to enhance student motivation, enjoyment, and performance. *Measurement in Physical Education and Exercise Science*, 12(3), 167–183. <http://doi.org/10.1080/10913670802216148>
- Yuan, K. H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, 40, 115-148. http://dx.doi.org/10.1207/s15327906mbr4001_5