Efectos del refuerzo intermitente sobre el mantenimiento de la condición física orientada a la salud en la educación física escolar

Effects of the intermittent reinforcement on health-related physical fitness maintenance in physical education

Daniel Mayorga Vega

Departamento de Educación Física y Deportiva
Facultad de Ciencias del Deporte
Universidad de Granada

2015

TESIS DOCTORAL INTERNACIONAL/ INTERNATIONAL DOCTORAL THESIS

PROGRAMA OFICIAL DE DOCTORADO EN CIENCIAS DE LA EDUCACIÓN

Editor: Universidad de Granada. Tesis Doctorales
Autor: Daniel Mayorga Vega
ISBN: 978-84-9125-525-3
URI: http://hdl.handle.net/10481/42438

TESIS DOCTORAL INTERNACIONAL/ INTERNATIONAL DOCTORAL THESIS

PROGRAMA OFICIAL DE DOCTORADO EN CIENCIAS DE LA EDUCACIÓN

UNIVERSIDAD DE GRANADA

EFECTOS DEL REFUERZO INTERMITENTE SOBRE EL MANTENIMIENTO DE LA CONDICIÓN FÍSICA ORIENTADA A LA SALUD EN LA EDUCACIÓN FÍSICA ESCOLAR

EFFECTS OF THE INTERMITTENT REINFORCEMENT ON HEALTH-RELATED PHYSICAL FITNESS MAINTENANCE IN PHYSICAL EDUCATION

DEPARTAMENTO DE EDUCACIÓN FÍSICA Y DEPORTIVA

FACULTAD DE CIENCIAS DEL DEPORTE

UNIVERSIDAD DE GRANADA

DANIEL MAYORGA VEGA

2015

A mis padres, hermana y Ania…

To my parents, sister and Ania…

DEPARTAMENTO DE EDUCACIÓN FÍSICA Y DEPORTIVA
FACULTAD DE CIENCIAS DEL DEPORTE
UNIVERSIDAD DE GRANADA

EFECTOS DEL REFUERZO INTERMITENTE SOBRE EL MANTENIMIENTO DE LA CONDICIÓN FÍSICA ORIENTADA A LA SALUD EN LA EDUCACIÓN FÍSICA ESCOLAR

EFFECTS OF THE INTERMITTENT REINFORCEMENT ON HEALTH-RELATED PHYSICAL FITNESS MAINTENANCE IN PHYSICAL EDUCATION

DANIEL MAYORGA VEGA

Director de Tesis [Thesis Supervisor]
Dr. Jesús Viciana Ramírez
Profesor Titular de Universidad
Universidad de Granada
España

Miembros del Tribunal [Committee]

Dr. Eduardo Cervelló Gimeno
Catedrático de Universidad
Universidad Miguel Hernández de Elche
España

Dra. María Belén Feriche Fernández-Castanys
Profesora Titular de Universidad
Universidad de Granada
España

Dr. Jarmo Liukkonen
Profesor Titular de Universidad
Universidad de Jyväskylä
Finlandia

Dr. Fernando del Villar Álvarez
Catedrático de Universidad
Universidad de Extremadura
España

Dr. Luis Javier Chirosa
Profesor Titular de Universidad
Universidad de Granada
España

En Granada, 5 de noviembre de 2015

Prof. Dr. Jesús Viciana Ramírez
Profesor Titular de Universidad

Departamento de Educación Física y Deportiva
Facultad de Ciencias del Deporte
Universidad de Granada

JESÚS VICIANA RAMÍREZ, PROFESOR TITULAR DE UNIVERSIDAD DE LA UNIVERSIDA DE GRANADA

CERTIFICA:

Que la Tesis Doctoral titulada "Efectos del refuerzo intermitente sobre el mantenimiento de la condición física orientada a la salud en la educación física escolar" que presenta D. **DANIEL MAYORGA VEGA** al superior juicio del Tribunal que designe la Universidad de Granada, ha sido realizada bajo mi dirección durante los años 2011-2015, siendo un excelente trabajo de investigación en una nueva línea de investigación abierta por el grupo HUM-764 y que guarda el rigor científico sobradamente para merecer el Título de Doctor, siempre y cuando así lo considere el citado Tribunal.

Fdo. Jesús Viciana Ramírez

En Granada, 5 de noviembre de 2015

El doctorando D. DANIEL MAYORGA VEGA y el director de la tesis Dr. JESÚS VICIANA RAMÍREZ:

Garantizamos, al firmar esta Tesis Doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Director de la Tesis

Fdo. Jesús Viciana Ramírez

Doctorando

Daniel M.V.

Fdo. Daniel Mayorga Vega

En Granada, 5 de noviembre de 2015

ÍNDICE DE CONTENIDOS [INDEX OF CONTENTS]

PROYECTOS DE INVESTIGACIÓN

[RESEARCH PROJECTS]

PROYECTOS DE INVESTIGACIÓN [RESEARCH PROJECTS]

El trabajo desarrollado y las publicaciones que componen la presente memoria de Tesis Doctoral están basados en los siguientes proyectos de investigación:

- Efectos del refuerzo intermitente sobre el mantenimiento de la condición física orientada a la salud en la educación física escolar (II) (AP2010-5905). Ministerio de Educación, Cultura y Deporte, Gobierno de España.
  Fecha: Desde el 1/05/2012 hasta el 29/02/2016
- Efectos del refuerzo intermitente sobre el mantenimiento de la condición física orientada a la salud en la educación física escolar (I). Vicerrectorado de Política Científica e Investigación, Universidad de Granada.
  Fecha: Desde el 1/03/2012 hasta el 30/04/2012
- Efecto de dos tipos de refuerzo intermitente sobre la salud tras veinte sesiones de de trabajo en secundaria. Ministerio de Educación, Gobierno de España.
  Fecha: Curso académico 2009/2010
- Efecto de dos tipos de refuerzo intermitente sobre la pérdida de la fuerza tras un periodo de trabajo en secundaria. Vicerrectorado de Política Científica e Investigación, Universidad de Granada.
  Fecha: Desde el 01/06/2009 hasta el 31/05/2010

BECAS DE INVESTIGACIÓN

[RESEARCH GRANTS]

BECAS DE INVESTIGACIÓN [RESEARCH GRANTS]

La presente memoria de Tesis Doctoral ha sido posible gracias a las siguientes becas y subvenciones:

- Ayuda de movilidad para estancias breves en el extranjero a beneficiarios del subprograma de Formación del Profesorado Universitario (FPU). *Józef Pisudski University of Physical Education in Warsaw*, Polonia. Financiado por el Ministerio de Educación, Cultura y Deporte, Gobierno de España.
  Fecha: Desde el 13/07/2015 hasta el 14/09/2015
- Ayuda de estancia de doctorado en el marco del Programa PAP-Erasmus. *Józef Pisudski University of Physical Education in Warsaw*, Polonia. Financiado por la Agencia Nacional Erasmus (OAPEE), Comisión Europea; Consejería de Educación, Junta de Andalucía; Ministerio de Educación, Cultura y Deporte, Gobierno de España.
  Fecha: Desde el 14/06/2013 hasta el 14/09/2013
- Ayuda del Programa de Formación del Profesorado Universitario (FPU) (AP2010-5905). Financiado por el Ministerio de Educación, Cultura y Deporte, Gobierno de España.
  Fecha: Del 1/05/2012 hasta el 29/02/2016
- Beca Formación de Investigadores, Programa 6A Beca-Contrato FPU. Financiado por el Vicerrectorado de Política Científica e Investigación, Universidad de Granada.
  Fecha: Del 1/03/2012 hasta el 30/04/2012
- Beca de Colaboración. Financiado por el Ministerio de Educación, Gobierno de España.
  Fecha: Curso académico 2009/2010
- Beca de Iniciación a la Investigación. Financiado por el Vicerrectorado de Política Científica e Investigación, Universidad de Granada.
  Fecha: Del 01/06/2009 hasta el 31/05/2010

PUBLICACIONES

[PUBLICATIONS]

PUBLICACIONES [PUBLICATIONS]

La presente memoria de Tesis Doctoral está compuesta por las siguientes publicaciones:

I. Viciana, J., & **Mayorga-Vega, D.** (2013). Análisis del cambio curricular de Educación Física en Primaria [Analysis of primary school curricular change from LOGSE to LOE in physical education]. *Profesorado*, *17*(3), 257-271. SJR: 0.100; Education: Q4.

II. Viciana, J., **Mayorga-Vega, D.**, & Mompeán Campillo, M. Cumplimiento de los estándares curriculares de condición física-salud en educación física. Estudio de la planificación en la formación inicial [Compliance of curriculum standards of health-related physical fitness in physical education. Study of the planning in pre-service teachers]. Pending acceptance of the revised version in *Cultura, Ciencia y Deporte.* SJR: 0.143; Sports Science: Q4.

III. Viciana, J., & **Mayorga-Vega, D.** Innovative teaching units applied to physical education. Changing the curriculum management for authentic outcomes. Accepted for publication in *Kinesiology.* JCR: 0.585; Sport Sciences: Q4.

IV. Viciana, J., **Mayorga-Vega, D.**, & Cocca, A. (2014). Modelo de aprendizaje exitoso en educación física y su mantenimiento. Estudio del efecto del refuerzo intermitente sobre la condición física. [Successful learning model in physical education and its maintenance. The intermittent reinforcement applied to physical fitness]. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, *9*(1), 155-171. SJR: 0.324; Sports Science: Q3.

V. **Mayorga-Vega, D.**, Aguilar-Soto, P., & Viciana, J. (2015). Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *Journal of Sports Science and Medicine*, *14*(3), 536-547. JCR: 1.025; Sport Sciences: Q3.

VI. **Mayorga-Vega, D.**, Bocanegra-Parrilla, R., Ornelas, M., & Viciana, J. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: A systematic review and meta-analysis. Submitted to *Plos One*. JCR: 3.234; Multidisciplinary Sciences: Q1.

VII. **Mayorga-Vega, D.**, Merino-Marban, R., & Viciana, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine*, *13*(1), 1-14. JCR: 1.025; Sport Sciences: Q3.

VIII. **Mayorga-Vega, D.**, Viciana, J., Cocca, A., & Merino-Marban, R. (2014). Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis. *Journal of Human Sport & Exercise*, *9*(1), 188-200. doi:10.4100/jhse.2014.91.18. SJR: 0.241; Physical Therapy, Sports Therapy and Rehabilitation: Q3.

IX. **Mayorga-Vega, D.**, & Viciana, J. (2015). Las clases de educación física solo mejoran la capacidad cardiorespiratoria de los alumnos con menor condición física: Un estudio de intervención controlado [Physical education classes only improve cardiorespiratory fitness of students with lower physical fitness: A controlled intervention study]. *Nutrición Hospitalaria*, *32*(1), 330-335. doi: 10.3305/nh.2015.32.1.8919. JCR: 1.040; Nutrition & Dietetics: Q4.

X. **Mayorga-Vega, D.**, Merino-Marban, R., Real, J. & Viciana, J. (2015). A physical education-based stretching program performed once a week also improves hamstring extensibility in schoolchildren: A cluster-randomized controlled trial. *Nutrición Hospitalaria*, *32*(4), 1715-1721. doi: 10.3305/nh.2015.32.4.9302. JCR: 1.040; Nutrition & Dietetics: Q4.

XI. **Mayorga-Vega, D.**, Merino-Marban, R., Redondo-Martín, F. J., & Viciana, J. Effect of a one-session-per-week physical education-based stretching program on hamstring extensibility in schoolchildren. Submitted to *Kinesiology.* JCR: 0.585; Sport Sciences: Q4.

XII. **Mayorga-Vega, D.**, Merino-Marban, R., Vera-Estrada, F., & Viciana, J. (2014). Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology*, *46*(2), 227-233. JCR: 0.585; Sport Sciences: Q4.

XIII. Merino-Marban, R., **Mayorga-Vega, D.**, Fernandez-Rodriguez, E., Vera, F., & Viciana, J. (2015). Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *European Physical Education Review*, *21*(1), 83-92. doi: 0.1177/1356336X14550942. JCR: 0.673; Education & Educational Research: Q3.

XIV. **Mayorga-Vega, D.**, Viciana, J., & Cocca, A. (2013). Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics*, *37*, 153-160. doi: 10.2478/hukin-2013-0036. JCR: 0.698; Sport Sciences: Q4.

XV. **Mayorga-Vega, D.**, Montoro-Escaño, J., Merino-Marban, R., & Viciana, J. (In press). Effects of a physical education-based program on health-related physical fitness and its maintenance in high school students: A cluster-randomized controlled trial. *European Physical Education Review.* doi: 10.1177/1356336X15599010. JCR: 0.673; Education & Educational Research: Q3.

XVI. **Mayorga-Vega, D.**, Merino-Marban, R., Manzano-Lagunas, J., Blanco, H., & Viciana, J. Effects of a stretching development and maintenance program on hamstring extensibility in schoolchildren: A cluster-randomized controlled trial. Pending acceptance of the revised version in *Journal of Sports Science and Medicine.* JCR: 1.025; Sport Sciences: Q3.

XVII. **Mayorga-Vega, D.**, Viciana, J., Cocca, A., & De Rueda Villén, B. (2012). Effect of a physical fitness program on physical self-concept and physical fitness elements in primary school students. *Perceptual and Motor Skills*, *115*(3), 984-996. doi: 10.2466/06.10.25.PMS.115.6. 984-996. JCR: 0.655; Experimental Psychology: Q4.

XVIII. **Mayorga-Vega, D.**, Montoro-Escaño, J., & Viciana, J. ¿La motivación autodeterminada hacia la educación física influye en la mejora de la condición física saludable? [Does the self-determination motivation toward physical education influence the physical fitness improving?]. Submitted to *Revista de Psicología del Deporte*. JCR: 0.487; Applied Psychology: Q4.

* JCR: Journal Citation Reports impact factor – ISI Web of Knowledge™/ SJR: SCImago Journal Rank impact factor – Scopus; Subject category: Quartile.

RESUMEN/

ABSTRACT

RESUMEN

Hoy en día la condición física es considerada como uno de los marcadores de salud y calidad de vida más potentes en la infancia y adolescencia. Lamentablemente, en la actualidad existen un gran número de niños con baja condición física. La asignatura de educación física podría desempeñar un papel muy importante ayudando a los escolares a alcanzar y mantener niveles saludables de condición física. Sin embargo, los profesores de educación física se enfrentan a diferentes limitaciones relacionadas con la planificación para alcanzar este objetivo. Por tanto, es necesario conocer los efectos empíricos de intervenciones en educación física que apoye la práctica basada en la evidencia de una planificación viable y eficaz para el desarrollo y mantenimiento de los niveles de condición física relacionados con la salud de los estudiantes.

El objetivo general de la presente Tesis Doctoral fue examinar el efecto de los programas de intervención de acondicionamiento físico en educación física sobre el desarrollo y mantenimiento de los niveles de condición física relacionada con la salud en escolares.

Los principales hallazgos de la presente Tesis Doctoral sugieren que: (a) La condición física relacionada con la salud tiene un papel muy importante en el currículo de la Ley Orgánica de Educación. (b) Los profesores en formación inicial aplican un número insuficiente de sesiones de intervención que imposibilitaría el incremento significativo de la condición física relacionada con la salud de los escolares. (c) Los tests de campo *classic sit-and-reach*, stand-*and-reach*, *20-m shuttle run*, *1.5 mile walk/run* y *12 min walk/run* presentan una adecuada validez de criterio en los niños. (d) Después de un programa de acondicionamiento física en educación física, sólo los estudiantes con menor nivel de condición física incrementan la condición física relacionada con la salud. (e) Un programa de estiramientos a largo plazo de una sesión semanal mejora la extensibilidad isquiosural de los estudiantes. (f) Después de un programa de desarrollo a corto plazo en educación física, un refuerzo intermitente que consiste en la realización de actividades deportivas de entrenamiento integrado durante sólo 10-15 minutos seguido por actividades de aprendizaje de estos deportes realizado dos sesiones por semana, no sólo es eficaz sino también viable para el mantenimiento de los niveles de fuerza muscular y capacidad cardiorespiratoria de los escolares. (g) Un refuerzo intermitente con dos sesiones semanales de estiramientos de sólo un minuto también es viable y eficaz en el mantenimiento de la extensibilidad isquiosural de los estudiantes en el contexto de la educación física escolar. (h)

La mejora en la condición física objetiva tras un programa de acondicionamiento físico en educación física no va acompañada de grandes cambios en el autoconcepto físico de los escolares. (i) Un programa de acondicionamiento físico en educación física solo mejora los niveles de capacidad cardiorrespiratoria de los estudiantes con una moderada-alta motivación autodeterminada hacia la educación física.

ABSTRACT

Physical fitness is currently considered as one of the most powerful markers of health and quality of life in childhood and adolescence. Unfortunately, nowadays low physical fitness affects a large number of children. The physical education subject may play an important role helping schoolchildren to achieve and maintain health-enhancing physical fitness levels. However, physical education teachers face several planning-related limitations to reach this purpose. Therefore, it is necessary to know the empirical effects of physical education-based interventions that support an evidence-based practice in planning a feasible and effective development and maintenance of students´ health-related physical fitness levels.

The overall objective of the present Doctoral Thesis was to examine the effect of the physical education-based development and maintenance intervention programs on health-related physical fitness levels in schoolchildren.

The main findings from the present Doctoral Thesis suggest that: (a) The health-related physical fitness has a very important role in the standard curriculum of the *Ley Orgánica de Educación*. (b) Pre-service teachers apply an insufficient number of intervention sessions that would make impossible the significant improvement of health-related physical fitness in schoolchildren. (c) The field tests classic sit-and-reach, stand-and-reach, 20-m shuttle run, 1.5 mile walk/run and 12 min walk/run show an adequate criterion-related validity among children. (d) After a physical education-based development program, only students with lower physical fitness levels improve health-related physical fitness. (e) A long-term stretching program performed once a week improves students´ hamstring extensibility. (f) After a short-term physical education-based development program, an intermittent reinforcement consisting of performing sports-integrated training activities for only 10-15 min followed by introduction activities to these sports two sessions per week, it is not only effective but also feasible for maintaining schoolchildren´s muscular and cardiorespiratory fitness levels. (g) An intermittent reinforcement with only one-minute stretching sessions twice a week is also feasible and effective in maintaining students´ hamstring extensibility in physical education setting. (h) The improvement in objective physical fitness through a physical education-based physical fitness program is not accompanied by major changes in schoolchildren´s physical self-concept. (i) A physical education-based physical fitness program only improves the cardiorespiratory fitness levels of students with a moderate-to-high self-determined motivation toward physical education.

ABBREVIATIONS

[ABREVIATURAS]

ABBREVIATIONS [ABREVIATURAS]

Abbreviation	**Definition**
ANCOVA	Analysis of covariance
ANEAE	*Alumnos con necesidades específicas de apoyo educativo*
ANOVA	Analysis of variance
BAH	Bent arm hang test
BMI	Body mass index
CFI	Comparative fit index
CF-S	*Condición física-salud*
CG	Control group
CI	Confidence interval
CV	Credibility interval
DE	*Desviación estándar*
EF	*Educación física*
EG	Experimental group
ESO	*Educación Secundaria Obligatoria*
ESR	20-m endurance shuttle run test
GC	*Grupo control*
GE	*Grupo experimental*
ICC	Intraclass correlation coefficient
LOA	Limits of agreement
LODE	*Ley Orgánica del Derecho a la Educación*
LOE	*Ley Orgánica de Educación*
LOGSE	*Ley de Ordenación General del Sistema Educativo*
LOMCE	*Ley Orgánica para la Mejora de la Calidad Educativa*
M	Mean
MANCOVA	Multivariate analysis of covariance
MDC	Minimal detectable change
PE	Physical education
PSDQ	Physical Self-Description Questionnaire
QUB	Queen´s University of Belfast
RMSEA	Root mean square error of approximation
SD	Standard deviation

SR	Sit-and-reach (test)
SUP	Sit-ups in 30 seconds test
TICs	*Tecnologías de la información y las comunicaciones*
TU	Teaching unit
TT	Toe-touch (test)
UD	*Unidad didáctica*
VO_2	Oxygen uptake
VO_2max	Maximal oxygen uptake
VO_2peak	Peak oxygen uptake
20MSR	20-m shuttle run (test)

INTRODUCTION

[INTRODUCCIÓN]

INTRODUCTION [INTRODUCCIÓN]

Physical fitness is currently considered as one of the most powerful markers of health and quality of life in childhood and adolescence (Ortega, Ruiz, Castillo, & Sjöström, 2008; Ruiz et al., 2009). For instance, there is strong evidence indicating that among children higher cardiorespiratory fitness levels are associated with a healthier cardiovascular profile in adulthood, and muscular strength improvements are negatively related to changes in overall adiposity (Ruiz et al., 2009). Although the number of related studies with flexibility is still limited, for instance, low hamstring extensibility has been associated to a higher risk of current low back pain (Feldman, Shrier, Rossignol, & Abenhaim, 2001; Jones, Stratton, Reilly, & Unnithan, 2005; Sjölie, 2004) and neck tension (Mikkelsson et al., 2006), as well as an increase in the risk of low back pain later in adulthood (Kujala, Taimela, Salminen, & Oksanen, 1994). Furthermore, in young people improvements in physical fitness have positive effects on psychological markers such as depression, anxiety, mood status and self-esteem, and seem also to be associated with a higher academic performance (Ortega et al., 2008).

Unfortunately, nowadays low physical fitness affects a large number of children. For instance, in Spain about one in five adolescents have a physical fitness level indicative of future cardiovascular risk (Ortega et al., 2005). Although physical fitness is in part determined genetically, it can also be significantly influenced by environmental factors such as physical activity or diet (Cuenca-García et al., 2014; Sygit et al., 2012). Therefore, health promotion policies should be designed to promote healthy behaviours such as encouraging children to achieve recommended levels of physical activity (Ortega et al., 2008). Since in most developed countries all the children attend school, especially through the physical education (PE) area, schools may play an important role identifying schoolchildren with low physical fitness levels, as well as helping them to achieve health-enhancing physical fitness levels (Ortega et al., 2008).

1. Compliance of curriculum alignment regarding health-related physical fitness in physical education

In recent decades in Spain the area of PE has undergone many important changes, especially by the major educational reforms (e.g., Jefatura del Estado, 1985, 1990, 2006, 2013). Any change in the educational law must be accompanied by practical changes such as, for instance, how teachers plan the PE sessions for their students (Viciana, Mayorga-Vega, & Merino-Marban, 2014). However, a superficial reading of the curriculum standards from the legislative documents leaves much uncertainty. Therefore, comparing the curriculums standards from the *Reales Decretos de Enseñanzas Mínimas para Educación Primaria y Secundaria* through a systematic and interpretive analysis of their content, as well as drawing practical conclusions for the PE teachers is necessary (Viciana, Salinas, & Cocca, 2007). This analysis would extract the evolutionary message of the national curriculum in order to provide practical guidance for PE teachers (Paper I).

The curricular area of PE has been considered one of the main strategies to ensure healthy physical fitness levels among schoolchildren (Ministerio de Educación, Cultura y Deporte, 2014, 2015). However, planning the improvement of students´ health-related physical fitness levels in the PE setting is a very complex task (Viciana et al., 2014). For instance, because of the need to deal with a large number of contents during the whole school year, PE teachers plan extremely shorts teaching units (TU), sometimes without meaning or effectiveness (Robles, Giménez, & Abad, 2010), leading to an inadequate development of the health-related physical fitness (Pérez-Pastur, 2010). Moreover, the TUs are usually designed as a particular number of lessons based on, in the best case scenario, in the own experience of the teacher and, in the worst of case scenario, in the mere intuition and even in the improvisation (Viciana & Zabala, 2004).

Therefore, the need of planning carefully the health-related physical fitness seems to be evident. For this purpose, following careful recommendations to ensure some success such as to assure the enough duration of intervention programs, to provide the necessary stimulus repetitions in order to prevent the physical fitness loss, or to follow the curricular standards guidelines of the educational stage and course are need (Viciana et al., 2014). Unfortunately, to our knowledge there are only a few qualitative case studies focused on the PE-based planning decisions (Emmer, 1986; Placek, 1984). However, these studies do not provide comprehensive conclusions that give us an idea if such planning has being done correctly. Therefore, making a deep diagnosis of the annual planning of PE teachers in order to analyze the effectiveness and compliance of the curriculum standards regarding the health-related physical fitness seems to be necessary (Paper II).

2. Physical education-based planning for developing and maintaining students' health-related physical fitness levels

Planning in education has been theoretically associated with the characteristics of flexibility, efficiency, and based on objectives (Viciana, 2002). In contrast to these planning characteristics, one of the most used planning structures in the subject of PE has been developed around the concept of TU (Kelly & Melograno, 2004), understood as a closed and traditional concept associated to several problems. Firstly, the concept of "covering the curriculum" (Siedentop & Tanehill, 2000), where teachers feel the necessity of treating a wide range of contents during a school year, has caused ineffective short TUs (Robles et al., 2010). For instance, PE teachers usually design very short TUs with the main objective that students improve their physical fitness levels (Pérez-Pastur, 2010). Secondly, TU has been conceived as a particular number of lessons consecutively developed in the school centre to attain an objective (or a group of them), but the necessity of maintaining the level of learning acquired is not often mentioned (Viciana et al., 2014). Finally, traditional TUs have usually been based on the achievement of isolated objectives (Kelly & Melograno, 2004; Siedentop & Tanehill, 2000), forgetting the relationships between those objectives and the rest of the subject, or the situational perspective of the learning (Smith, 2011).

From this traditional point of view, the most important aspect of the TU in PE has been the duration, which it is represented by the number of lessons (minutes) needed to achieve a certain objective (Viciana, 2002). This duration includes the learning time spent and the time needed to learn by students as a part of it (Van der Mars, 2006). However, based on the original conception of planning with the attributes of flexibility, efficiency and based on educational objectives, and on an innovative TU conception, a new variable could be taken into account for improving those models and the students´ learning: this is the *innovative distribution of the TUs time* throughout an annual planning in PE. Maintaining the time spent learning and the time needed to learn constant, this innovative distribution of the time of the TUs would improve three important elements regarding the learning: (a) to provide situational learning in PE; (b) to improve the relational cognitive-behavioural learning between PE contents; and (c) to develop and maintain the cognitive and behavioural learning achieved. Based on meaningful learning instead of being contented with easily forgettable simple goals, the new perspective of the time-learning distribution along the academic year could facilitate teachers to attain authentic outcomes (Paper III). Its specific application in key issues in PE such as the developing and maintaining of students´ health-related physical fitness levels seems necessary (Paper IV).

3. Criterion-related validity of field tests for estimating physical fitness among children

The choice of a physical fitness test must be based on its feasibility and validity. Validity which refers to the ability of a test to reflect what it is designed to measure, it is the most important measurement characteristic (Baumgartner, Jackson, Mahar, & Rowe, 2015). Physical fitness can be valid and accurately measured through laboratory tests. For instance, cardiorespiratory fitness is widely operationalized as the maximal oxygen uptake (VO_2max) attained during a graded maximal exercise test (Pescatello, Arena, Riebe, & Thompson, 2014). Particularly the VO_2max attained during a laboratory-based and graded maximal exercise test is commonly considered the criterion measure of cardiorespiratory fitness (Pescatello et al., 2014). Nevertheless, since laboratory testing requires sophisticated and expensive equipment, qualified examiners, and long testing sessions, this technique is not feasible in school setting (Meredith & Welk, 2010).

Unlike the laboratory tests to directly determinate physical fitness, in the school setting the performance score obtained during field tests could be a feasibility alternative to assess physical fitness among children (Castro-Piñero et al., 2010). Physical fitness field tests require minimal equipment, are easy to administer, and are not too time-consuming, allowing the evaluation of a large number of children in a short period of time. Since the early interest in youth physical fitness testing in the 1950s (Zhu, Mahar, Welk, Going, & Cureton, 2011), many field tests have been proposed for children (Castro-Piñero et al., 2010). However, in order for a physical fitness field test to be considered "valid", it should measure, or at least to be highly associated with, what it is supposed to measure (Baumgartner et al., 2015). Particularly, criterion-related validity refers to the extent to which a field test is associated with the criterion measure or "gold standard" (Baumgartner et al., 2015). Nevertheless, currently there is no still consensus regarding the most appropriate field tests for estimating physical fitness among children (Ruiz, Ortega, & Castro-Piñero, 2015).

Each primary study that is published about the criterion-related validity of the physical fitness field tests among children only constitutes a single piece of a constantly growing body of evidence (Cooper, Hedges, & Valentine, 2009). However, while in some studies the correlation coefficient is high (Chatterjee, Banerjee, & Majumdar, 2006), in others the association is moderate or even low (Von Haaren, Härtel, Seidel, Schlenker, & Bös, 2011). To make sense of the often conflicting results found in the scientific literature, meta-analyses should be conducted (Cooper et al., 2009; Schmidt & Hunter, 2015). Therefore, meta-analyses remain a useful tool for the evaluation of evidence (Cooper et al., 2009), forming a critical process for the development of theory in science (Schmidt & Hunter, 2015). Consequently, knowing the population mean of the criterion-related validity of the field tests would help scientists, practitioners and users to select the most feasible and valid tests for estimating physical fitness among children (Papers V-VIII).

4. Effectiveness of physical education-based intervention programs for developing and maintaining health-related physical fitness in schoolchildren

A PE-based planning problem for developing students' health-related physical fitness levels is the necessity to "deliver" a large volume of curricular content during each academic year (Viciana et al., 2014). Additionally, PE potential is usually restricted by its limited curriculum time allocation (Hardman, 2008), particularly when the weekly frequency of sessions is only twice a week, which is the norm in most European countries (European Commission/ EACEA/ Eurydice, 2013). Therefore, despite the fact that a long-term program is the best way to improve physical fitness (Meyer et al., 2014), in many countries the application of a short-term physical fitness development program is one feasible option in the PE setting (Viciana et al., 2014). In this line, previous PE-based studies found that a short-term intervention program performed only twice a week can improve students' physical fitness levels (e.g., Faigenbaum et al., 2015; Ramírez et al., 2012; Sánchez, Mayorga-Vega, Fernández, & Merino-Marban, 2014). However, the influence of students´ physical fitness baseline levels on the effectives of these PE-based intervention programs is not deeply known yet (Papers IX and XVI). Additionally, due to restricted curriculum time allocation in relation with the large volume of curricular contents, it would be practical for teachers to know if PE-based intervention programs with only one session per week could be also effective for improving health-related physical fitness (Papers X-XI).

Another PE-based planning problem related to physical fitness is its expected decrease after a period of detraining. For instance, previous studies have found that after 8 to 12 weeks of detraining, children show a significant loss in the obtained gains in muscular strength levels (e.g., Ingle, Sleap, & Tolfrey, 2006; Tsolakis, Vagenas, & Dessypris, 2004). However, the flexibility detraining effects among schoolchildren should be also examined

(Papers XII-XIII). Additionally, many PE teachers conceive planning as "watertight drawers" that they have to fill with curricular contents (Siedentop & Tanehill, 2000). Furthermore, besides the large volume of curricular contents in relation with the restricted curriculum time allocation, another PE-based planning limitation is the fact that the academic year is frequently interrupted by several holiday periods, excursions, and other organized educational activities (Viciana et al., 2014).

Therefore, PE teachers usually carry out some exercises in their classes only for a few weeks, and when they cease doing them, they do not concern themselves with how long the effect will last. In this line, Viciana et al. (2014) suggested that, after a physical fitness development program, PE teachers should include some maintenance programs in order to retain students' physical fitness levels throughout the entire academic year. Apart from maintaining the physical fitness levels previously obtained, these programs would not interfere in the normal teaching of other PE curricular contents (Viciana et al., 2014). Consequently, knowing the empirical effects of the PE-based development and maintenance programs on health-related physical fitness is need to support evidence-based planning (Papers XIV-XVI). Figure 1 proposes a theoretical model for the PE-based planning for developing and maintaining students' health-related physical fitness levels.

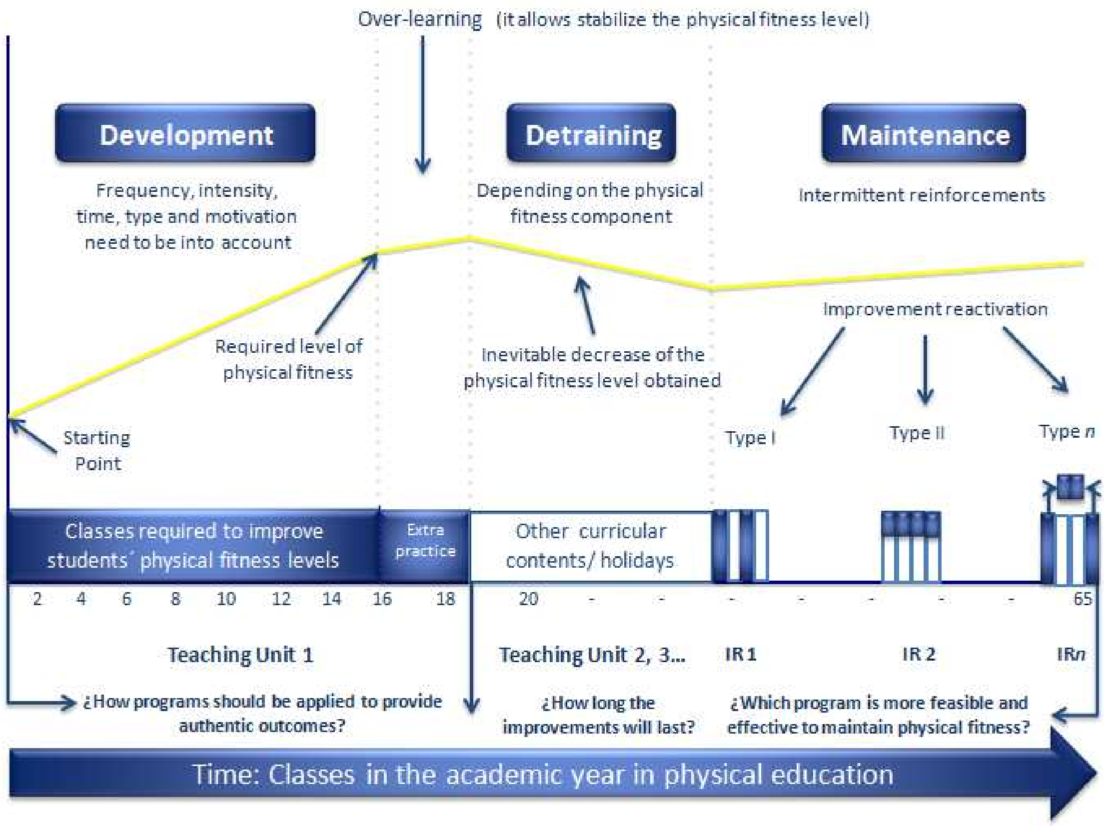


Figure 1. Successful model for the physical education-based planning to develop and maintain students' health-related physical fitness levels (modified from Viciana et al., 2014).

5. Effects of physical education-based physical fitness programs on physical self-concept in schoolchildren

Health is regarded not merely as the freedom from disease or injury, but also a state of complete physical, mental, and social well-being (World Health Organization, 1946). Childhood has been considered a crucial period of life, since dramatic physiological and psychological changes take place at these ages (Harter, 2012). For this reason, schools, mainly through the area of PE, have taken steps to make students aware of physical activity for their well-being demonstrating a responsible attitude towards their-selves and other people, and recognizing the healthy effects of physical activity (Ministerio de Educación, Cultura y Deporte, 2014, 2015).

Self-concept gains great importance during the childhood, since it is at this stage when individuals experience major changes regarding their skills and their ability to assess their skills (Harter, 2012). Specifically, physical self-concept, defined as the perception individuals have of their fitness, health, appearance, and physical activity (Tomás, Marsh, González-Romá, Valls, & Nagengast, 2014), is one of the most important elements of overall self-concept (Klesges et al., 1992). Physical self-concept is particularly relevant because of its impact on the levels of physical activity (Planinsec & Fosnaric, 2005) and physical fitness (Carraro, Scarpa, & Ventura, 2010; Mayorga, Viciana, & Cocca, 2012), with its corresponding influences on health (Ruiz et al., 2009), use of leisure time, and social relationships (Alfermann & Stoll, 2000).

Several previous studies have examined the effects of sport interventions on children´s physical self-concept (e.g., Salokun, 1994). However, studies examining the effects of physical fitness program on physical children´s physical self-concept are limited and contradictory (Faigenbaum et al., 1997; Greene & Ignico, 1995), especially in the PE setting (Sadres, Eliakim, Constantini, Lidor, & Falk, 2001). Because of the important role played by school-based programs promoting the development of physical and psychosocial health-related markers, a deep examination of the effects of PE-based physical fitness programs on physical self-concept is required (Papers XV and XVII).

6. Influence of motivation toward physical education on the effectiveness of intervention programs for improving students´ health-related physical fitness

As mentioned before, PE potential is limited by many factors such as its restricted curriculum time allocation, the large volume of curricular contents, or the constant interruption by several holiday periods, excursions, and other organized educational activities. Additionally, the success of the subject will also be limited if students are not motivated to actively participate in their PE sessions (Ntoumanis, 2001). Motivation is a psychological feature that arouses an individual to act toward a desired goal and elicits, controls, and sustains certain goal-directed behaviors; therefore, motivation is the purpose or psychological cause of any action (Ryan, Williams, Patrick, & Deci, 2009). In this line, numerous studies have found that the recommendation of spending at least 50% of PE sessions on moderate-to-vigorous physical activity rarely is met (Fairclough & Stratton, 2005, 2006).

Self-Determination Theory (Deci & Ryan, 2000) is a framework widely used to understand the consequences of motivation toward PE (Ntoumanis, 2001). This theory adopts a multidimensional perspective on motivation, suggesting that behavioral regulation toward an activity may be ordered on a *continuum* according to the extent to which motivation is self-determined (Deci & Ryan, 2000). Several previous studies have found how among schoolchildren self-determined motivation toward PE is positively associated to overall physical activity levels (e.g., Kalaja, Jaakkola, & Liukkonen, 2010; Ullrich-Francés & Cox, 2009; Yli-Piipari, Watt, Jaakkola, Liukkonen, & Nurmi, 2009). Similarly, the results of few studies have showed how students with higher self-determined motivation toward PE are also more physically active during their PE sessions than those with lower motivation (Lonsdale, Sabiston, Raedeke, Ha, & Sum, 2009; Mayorga-Vega & Viciana, 2014). Unfortunately, in spite of its great importance, to our knowledge there are not previous studies examining the influence of students´ motivation toward PE on the effectiveness of intervention programs for improving health-related physical fitness levels (Paper XVIII).

OBJETIVOS/

AIMS

OBJETIVOS

General

El objetivo general de la presente Tesis Doctoral fue examinar el efecto de los programas de intervención de acondicionamiento físico en educación física (EF) sobre el desarrollo y mantenimiento de los niveles de condición física relacionada con la salud en escolares.

Específicos

Los objetivos específicos de la presente Tesis Doctoral fueron los siguientes:

I. Comparar el contenido de los currículos de Reales Decretos de Enseñanzas Mínimas en Primaria de la Ley de Ordenación General del Sistema Educativo (LOGSE) y Ley Orgánica de Educación (LOE) mediante un análisis de contenido documental (Artículo I).

II. Analizar y detectar posibles deficiencias en la planificación de la condición física relacionada con la salud de los profesores de EF en formación inicial (Artículo II).

III. Proponer estructuras innovadoras de distribución del tiempo en la planificación de EF que permitan alcanzar resultados significativos, especialmente en el desarrollo y mantenimiento de la condición física relacionada con la salud (Artículos III-IV).

IV. Estimar y comparar el valor poblacional de la validez de criterio de diferentes tests de campo para estimar la condición física en personas aparentemente sanas; y examinar la influencia del sexo, edad y nivel de condición física de los individuos sobre los valores de validez de criterio de dichos tests (Artículos V-VIII).

V. Examinar la influencia de la condición física basal de los escolares en el efecto de un programa de intervención de acondicionamiento físico en EF sobre sus niveles de condición física relacionada con la salud (Artículos IX y XVI).

VI. Evaluar el efecto de un programa de estiramientos de una sesión semanal en las clases de EF sobre la extensibilidad isquiosural en escolares; y comparar los efectos de un programa de estiramientos realizado una y dos veces por semana en las clases de EF sobre la extensibilidad isquiosural en escolares (Artículos X-XI).

VII. Valorar el efecto de un periodo de desentrenamiento de la flexibilidad posterior a un programa de estiramientos a corto plazo en EF sobre la extensibilidad isquiosural en escolares (Artículos XII-XIII).

VIII. Estudiar el efecto de un programa de desarrollo a corto plazo en EF seguido por un programa de mantenimiento (es decir, refuerzo intermitente) con unidades didácticas (UD) alternadas e intermitentes sobre la mejora y mantenimiento de los niveles objetivos de condición física relacionada con la salud en escolares (Artículos XIV-XVI).

IX. Investigar el efecto de un programa de intervención de acondicionamiento físico en EF sobre los niveles de autoconcepto físico y condición física percibida de los escolares (Artículos XV y XVII).

X. Examinar la influencia de la motivación hacia la EF de los escolares en el efecto de un programa de intervención de acondicionamiento físico en EF sobre sus niveles de capacidad cardiorespiratoria (Artículo XVIII).

AIMS

Overall

The overall objective of the present Doctoral Thesis was to examine the effect of PE-based development and maintenance intervention programs on health-related physical fitness levels in schoolchildren.

Specifics

The specific objectives of the present Doctoral Thesis were the following:

I. To compare the content of the curriculum standards of Royal Decrees of Minimum Teaching in Primary Education of the *Ley de Ordenación General del Sistema Educativo* (LOGSE) and *Ley Orgánica de Educación* (LOE) through a documental analysis of content (Paper I).

II. To analyze and identify potential deficiencies in the planning of health-related physical fitness of pre-service PE teachers (Paper II).

III. To propose innovative structures of the time distribution in PE planning that allow to achieve authentic outcomes, especially in the development and maintenance of health-related physical fitness (Papers III-IV).

IV. To estimate and compare the population mean of the criterion-related validity coefficients of several field tests for estimating hamstring extensibility and cardiorespiratory fitness among apparently healthy people; and to examine the influence of individuals´ sex, age, and physical fitness level in criterion-related validity coefficients of these tests (Papers V-VIII).

V. To examine the influence of students´ physical fitness baseline on the effect of a PE-based physical fitness intervention program on their health-related physical fitness levels (Papers IX and XVI).

VI. To evaluate the effect of a PE-based stretching intervention program performed once a week on hamstring extensibility in schoolchildren; and to compare the effects of a PE-based stretching program performed once and twice a week on hamstring extensibility in schoolchildren (Papers X-XI).

VII. To assess the effect of a flexibility detraining period after a short-term PE-based stretching intervention program on hamstring extensibility in schoolchildren (Papers XII-XIII).

VIII. To examine the effect of a short-term PE-based development program followed by a maintenance program (i.e., intermittent reinforcement) with alternated and intermittent TUs on the improvement and maintenance of objective health-related physical fitness levels in schoolchildren (Papers XIV-XVI).

IX. To study the effect of a PE-based physical fitness intervention program on schoolchildren´s physical self-concept and perceived physical fitness levels (Papers XV and XVII).

X. To examine the influence of students´ motivation toward PE on the effect of a PE-based physical fitness intervention program on their cardiorespiratory fitness levels (Paper XVIII).

MATERIAL AND METHODS

[MATERIAL Y MÉTODOS]

MATERIAL AND METHODS [MATERIAL Y MÉTODOS]

The material and methods section of the present Doctoral Thesis is summarized in the Table 1. This table includes the most relevant methodological features from the scientific studies that compose the present Doctoral Thesis. For further information of any study, please check the material and methods sections of the corresponding papers.

Table 1. Summary of the main methodological features of the studies of the present Doctoral Thesis[a].

Paper	Study design	Participants	Procedure	Measures	Statistical analysis
I. Analysis of primary school curricular change from LOGSE to LOE in physical education	Descriptive	Curriculum standards Royal Decrees of Minimum Teaching in Primary Education of the LOGSE and LOE	Systematic content analysis	42 subject categories in 5 dimensions	Frequencies and percentages
II. Compliance of curriculum standards of health-related physical fitness in physical education. Study of the planning in pre-service teachers	Cross-sectional	180 PE planning from pre-service teachers of 19-38 years (146 males and 34 females)	Systematic content analysis	31 categories	Frequencies and percentages
V. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis	Systematic review and meta-analysis	57 studies included in the meta-analysis (78 correlations)	12 bibliographic databases and 5 additional modes of searching	15 categories including criterion-related validity coefficients	Hunter-Schmidt's psychometric meta-analysis
VI. Criterion-related validity of the distance- and time-based walk/run field tests for estimating cardiorespiratory fitness: A systematic review and meta-analysis	Systematic review and meta-analysis	123 studies included in the meta-analysis (201 correlations)	7 bibliographic databases and 5 additional modes of searching	16 categories including criterion-related validity coefficients	Hunter-Schmidt's psychometric meta-analysis
VII. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis	Systematic review and meta-analysis	34 studies included in the meta-analysis (150 correlations)	7 bibliographic databases and reference lists searching	11 categories including criterion-related validity coefficients	Hunter-Schmidt's psychometric meta-analysis
VIII. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis	Systematic review and meta-analysis	6 studies included in the meta-analysis (12 correlations)	5 bibliographic databases and reference lists searching	8 categories including criterion-related validity coefficients	Hunter-Schmidt's psychometric meta-analysis

Table 1. Continued.

IX. Physical education classes only improve cardiorespiratory fitness of students with lower physical fitness: A controlled intervention study	Cluster-randomized controlled trial and *ex post facto*	71 sixth-grade primary students (31 boys and 39 girls) 107 second-grade secondary students (68 boys and 39 girls)	EG1-2: 8 weeks, 2 sessions/week, 50 min EG1-2: 9 weeks, 2 sessions/week, 50 min	20-m shuttle run test	One-way ANOVA including *group* as fixed factor and *change* as dependent variable
X. A physical education-based stretching program performed once a week also improves hamstring extensibility in schoolchildren: A cluster-randomized controlled trial	Cluster-randomized controlled trial balanced by grade	163 first/second-grade secondary students (84 boys and 79 girls)	EG1: 8 weeks, 1 session/week, 4 min/session EG2: 8 weeks, 2 sessions/week, 4 min/session	Classic sit-and-reach test	One-way ANOVA including *group* as fixed factor and *change* as dependent variable
XI. Effect of a one-session-per-week physical education-based stretching program on hamstring extensibility in schoolchildren	Cluster-randomized controlled trial	37 third-grade primary students (18 boys and 19 girls)	EG: 32 weeks, 1 session/week, 3 min/session	Classic sit-and-reach test	Two-way ANOVA including *group* as independent variable and *time* as dependent variable
XII. Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren	Cluster-randomized controlled trial	45 sixth-grade primary students (26 boys and 19 girls)	EG: Development: 8 weeks, 2 sessions/week, 6 min/session Detraining: 5 weeks	Classic sit-and-reach test	Two-way ANOVA including *group* as independent variable and *time* as dependent variable

Table 1. Continued.

XIII. Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren	Cluster-randomized controlled trial	45 first-grade primary students (26 boys and 19 girls)	EG: Development: 8 weeks, 2 sessions/week, 1 min/session Detraining: 5 weeks	Classic sit-and-reach test	Two-way ANOVA including *group* as independent variable and *time* as dependent variable
XIV. Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren.	Cluster-randomized controlled trial	72 sixth-grade primary students (40 boys and 32 girls)	EG: Development: 8 weeks, 2 sessions/week, 50 min Detraining: 4 weeks Maintaining: 4 weeks, 1 session/week, 50 min	Sit-ups in 30 s test Bent arm hang test 20-m shuttle run test	Two-way ANOVA including *group* as independent variable and *time* as dependent variable
XV. Effects of a physical education-based program on health-related physical fitness and its maintenance in high school students: A cluster-randomized controlled trial	Cluster-randomized controlled trial	111 second-grade secondary students (70 boys and 41 girls)	EG: Development: 9 weeks, 2 sessions/week, 50 min Detraining: 4 weeks Maintaining: 8 weeks, 2 sessions/week, 10-15 min	International fitness scale Contour drawing rating scale Flexion-extension legs test 20-m shuttle run test	Two-way ANOVA/ ANCOVA including *group* as independent variable and *time* as dependent variable

Table 1. Continued.

XVI. Effects of a stretching development and maintenance program on hamstring extensibility in schoolchildren: A cluster-randomized controlled trial	Cluster-randomized controlled trial balanced by grade	140 third/fourth-grade primary students (66 boys and 74 girls)	EG1-2: Development: 9 weeks, 2 sessions/week, 4 min/session EG1-2: Detraining: 5 weeks EG1-2: Maintaining: 11 weeks, 2 sessions/week; EG1: 4 min/session, EG2: 1 min/session	Classic sit-and-reach test	Two-way ANOVA including *group* as independent variable and *time* as dependent variable
XVII. Effect of a physical fitness program on physical self-concept and physical fitness elements in primary school students	Cluster-randomized controlled trial	75 sixth-grade primary students (41 boys and 34 girls)	EG: 8 weeks, 2 sessions/week, 50 min	PSDQ Bent arm hang test Sit-ups in 30 s test Standing long jump test 20-m shuttle run test	Two-way MANCOVA including *group* as independent variable, *time* as dependent variable and *pre-intervention* as covariable

Table 1. Continued.

XVIII. Does the self-determination motivation toward physical education influence the physical fitness improving?	Cluster-randomized controlled trial and *ex post facto*	97 second-grade secondary students (62 boys and 35 girls)	EG1-3: 22 weeks, 2 sessions/week, 50 min	20-m shuttle run test	One-way ANCOVA including *group* as fixed factor, *change* as dependent variable and *pre-intervention* as covariable

Note. LOGSE: *Ley Orgánica General del Sistema Educativo*; LOE: *Ley Orgánica de Educación*; PE: Physical Education; EG: Experimental group; ANOVA: Analysis of variance; ANCOVA: Analysis of covariance; PSDQ: Physical Self-Description Questionnaire; MANCOVA: Multivariate analysis of covariance. [a] Papers III and IV are not reported because they are theoretical studies.

RESULTS AND DISCUSSION

[RESULTADOS Y DISCUSIÓN]

RESULTS AND DISCUSSION [RESULTADOS Y DISCUSIÓN]

The results and discussion section of the present Doctoral Thesis is shown as a compilation of scientific papers. They are enclosed in the form that have been published or submitted.

1. COMPLIANCE OF CURRICULUM ALIGNMENT REGARDING HEALTH-RELATED PHYSICAL FITNESS IN PHYSICAL EDUCATION (PAPERS I-II)

Paper I

ANÁLISIS DEL CAMBIO CURRICULAR DE EDUCACIÓN FÍSICA EN PRIMARIA

[ANALYSIS OF PRIMARY SCHOOL CURRICULAR CHANGE FROM LOGSE TO LOE IN PHYSICAL EDUCATION]

Viciana, J., & Mayorga-Vega, D.

Profesorado

2013, *17*(3), 257-271.

Profesorado
Revista de currículum y formación del profesorado

VOL. 17, Nº 3 (sept.-diciembre 2013)
ISSN 1138-414X (edición papel)
ISSN 1989-639X (edición electrónica)
Fecha de recepción 08/03/2013
Fecha de aceptación 05/07/2013

ANÁLISIS DEL CAMBIO CURRICULAR DE LOGSE A LOE EN LA EDUCACIÓN FÍSICA DE PRIMARIA

Analysis of Primary School Curricular Change from LOGSE to LOE in Physical Education

Jesús ***Viciana Ramírez*** *y Daniel* ***Mayorga-Vega***
Universidad de Granada
E-mail: jviciana@ugr.es , dmayorgavega@gmail.com

Resumen:
En ocasiones, el profesorado tiene problemas para la implantación de las Reformas educativas al aula, normalmente por la interpretación de los cambios propuestos por la Administración en el área curricular correspondiente. Por ello, el objetivo fue analizar comparativamente, siguiendo a Viciana, Salinas y Cocca (2007), el contenido de los Reales Decretos de Enseñanzas Mínimas entre la Ley de Ordenación General del Sistema Educativo y la nueva Ley Orgánica de Educación en la etapa de Primaria, con el fin de esclarecer los cambios propuestos y ser aplicados al aula por el profesorado más fácilmente. Se utilizó el análisis de contenido de los textos como metodología de estudio, reduciendo los datos a categorías temáticas de manera inductiva y realizando un conteo de frecuencias de aparición de dichas categorías en ambos textos. Se indujeron cinco dimensiones con 42 categorías detallando el cambio en cada una de ellas. Los resultados mostraron que hay notables variaciones en el contenido de dichos Reales Decretos, reflejando el cambio de los intereses sociales de una Ley a otra. Igualmente dedujimos hacia dónde deben encaminar su docencia los profesores de EF según la importancia otorgada en cada documento legislativo a las categorías definidas, haciéndoles reflexionar críticamente sobre estos cambios.
Palabras clave: ***Reforma educativa, legislación educativa, currículum, análisis de contenido, educación Primaria***

Abstract:

Educational changes are carried out with the objective to adapt the curriculum to current social and cultural needs. Teachers need to implement these educational changes into their classes, but they usually to encounter difficulties in the interpretation of the national curriculum. A content analysis comparing the frequencies of categories that appear in both past and present legislative documents is presented, following the original analysis made by Viciana, Salinas, and Cocca, (2007). A practical and understandable perspective of educational change by showing the new educational tendencies was provided. This led us to create five dimensions with 42 categories in which we detail the change occurring in each. Results show that there are substantial modifications in the content of both educational legislations, reflecting the social change from one to the other. This lets us guide all physical education teachers in the planning of their teaching at a Primary School level and prompt their critical thinking about these changes.

Key words: ***Educational change, educational legislation, national curriculum, content analysis primary School***

1. Introducción

Desde incluso la época anterior a la Ley Orgánica del Derecho a la Educación (LODE, 1985), la Educación Física (EF) ha sufrido cambios muy importantes hasta la actualidad. Los principales cambios han venido de la mano de las grandes Reformas Educativas. La incorporación del alumnado como centro de la enseñanza y su actividad en el proceso de enseñanza; la reforma de los métodos pedagógicos para impartir la docencia en función de dicho alumnado; el papel de los profesores, tanto a nivel docente como social en el aula; la derivación de competencias con los niveles del currículo, su apertura y el protagonismo de los centros y profesores en su aplicación; la innovación y la investigación como parte de la función docente; la educación por competencias, etc., han sido ejemplos de estos grandes cambios. Sin embargo, acompañando a estos cambios y en buena parte producidos por ellos (Viciana, 2000), han aparecido nuevas corrientes, tendencias innovadoras y enfoques de la EF que han hecho evolucionar nuestro área curricular en todo este tiempo.

La EF ha evolucionado de una manera pendular desde sus inicios como parte del currículo escolar estructurado hasta la actualidad. En ambos extremos del péndulo se situarían los enfoques de proceso y de producto de los aprendizajes, habiéndose experimentado vaivenes en todo este tiempo (Viciana, 2002). Nos acordamos por ejemplo de la gimnasia sueca, casi militarista, de Pier Henrich Ling [1776-1839] de movimientos analíticos y diferenciados que caracterizó a la época franquista; del enfoque de producto, de rendimiento y de deportes tradicionales que caracterizó a la época de la LODE; del enfoque lúdico, vivencial y de deportes alternativos de la Ley de Ordenación General del Sistema Educativo (LOGSE, 1990); y de la educación por competencias y la EF enfocada hacia la salud como principales enfoques actuales de la Ley Orgánica de Educación (LOE, 2006). Pero al margen de las grandes corrientes educativas de la EF, ¿cómo debemos enfocar la EF en las escuelas actualmente? ¿En qué difieren la LOGSE y la actual LOE para los maestros especialistas y los profesores de EF?, y ¿Cómo deben cambiar su docencia de la EF hacia los contenidos y objetivos principales que nuestra sociedad demanda?

Varios han sido los análisis realizados sobre los cambios ocurridos, sin embargo, estos análisis se han centrado en las aportaciones más novedosas y destacables, como el realizado por Learreta (2003) entre la LOGSE y la nunca implantada Ley Orgánica de Calidad de la Educación (LOCE). En otras ocasiones, se ha realizado un análisis más pormenorizado del primer nivel curricular por cuanto abarcan a los apartados más importantes del mismo: fines, objetivos, contenidos, criterios de evaluación y orientaciones metodológicas (Herrador, Huertas y Lara, 2009; Muñoz, 2007).

En definitiva, todo cambio general en la Ley de Educación debe acompañarse de cambios prácticos en la forma de actuar en clase, en la forma de planificar la EF para los alumnos, en la manera de redactar los objetivos y las intenciones de la enseñanza, en los contenidos, los enfoques metodológicos que se emplean para su enseñanza-aprendizaje, etc. Sin embargo, la lectura de los primeros niveles del currículo en los documentos legislativos deja bastante incertidumbre si sólo les hacemos una lectura superficial, ya que se mencionan en ellos muchos temas de interés y se atiborra de contenidos específicos y de orientaciones metodológicas al docente que las lee.

Nuestro objetivo principal es, por tanto, comparar estos dos currículos (Reales Decretos de Enseñanzas Mínimas en Primaria de LOGSE y LOE) con un análisis sistemático e interpretativo del contenido, extrayendo conclusiones para la EF. Realizamos esto a partir de un conteo de frecuencias de núcleos y centros de interés que se hacen explícitos y se deducen de los documentos analizados. Con este sistema estructurado de análisis intentamos extraer el mensaje evolutivo del primer nivel de concreción del currículo para aportar directrices prácticas a todos los profesionales que deben adaptar sus enseñanzas a esta nueva filosofía de enseñanza. Igualmente, aportamos nuestros comentarios particulares a modo de discusión de resultados, opinando sobre los hallazgos resultantes del análisis categorial y apoyándonos en la literatura previa relacionada.

2. Método

Para el propósito de estudio de comparar documentos escritos, aplicar una metodología de análisis contrastada y científica basada en dichos documentos publicados en los Boletines Oficiales del Estado (Reales Decretos 1006/1991 y 1513/2006 de Enseñanzas Mínimas para Primaria en la LOGSE y LOE, respectivamente) era lo más adecuado. Para ello seguimos la metodología de análisis documental propuesta por Viciana, Salinas y Cocca (2007), basada en la propuesta de Miles y Huberman (1984), que combina la investigación cualitativa y cuantitativa para extraer sistemáticamente los mensajes ocultos en el texto.

Para evitar falsas interpretaciones surgidas de una lectura superficial, a este análisis cualitativo e interpretativo se le añade la sistematicidad y la cuantificación de las apariciones temáticas en los documentos (Miles y Huberman, 1984; Goetz y Le Compte, 1984). Así, obtuvimos consensuadamente entre los investigadores, un listado de categorías temáticas que se aplicaba al texto mediante un conteo de frecuencias en los documentos comparados. El último paso consistió en deducir de la importancia de estas frecuencias para ambas leyes, el análisis y la evolución social de dichos temas, orientando a los docentes sobre cómo actuar en clase de EF, es decir, concluyendo desde la comparación realizada nuevas metodologías de trabajo en el aula que favorecieran la puesta en práctica de los cambios implícitos en la Ley.

En un primer momento se leyeron los documentos y se extrajeron las categorías temáticas indiscriminadamente. Posteriormente se discutieron en el seno de un grupo de cinco expertos (dos licenciados y tres doctores en Educación Física que habían realizado su tesis doctoral usando esta metodología de análisis cualitativa) para estructurar un sistema de categorías adecuado (redefiniendo dichas categorías, modificándolas y agrupándolas en dimensiones temáticas según su contenido y las relaciones entre ellos). Y finalmente se procedió a una lectura pormenorizada de todo el texto, realizando la codificación y el conteo sistemático de apariciones de dichas categorías en ambos textos legislativos. Las tablas de resultados y su interpretación nos permiten proponer las líneas principales de cambio deducidas de los documentos analizados. Estos pasos metodológicos nos posibilitan deducir mensajes para los docentes y transmitirles los principales cambios ocurridos en el currículo de EF, evitando la confusión inicial que produce la lectura comparativa de ambos documentos. La exposición en detalle de las fases de inducción de categorías la podemos ver en Viciana y Sánchez (2002).

El sistema de categorías temáticas consistió en cinco dimensiones que englobaban un total de 42 categorías, estructuradas de la siguiente manera:

1. Habilidades y destrezas, condición física y desarrollo motor (con nueve categorías). Esta dimensión trata fundamentalmente del aprendizaje y el desarrollo motor en los niños. En ella se encuadraron categorías relacionadas con las habilidades y el aprendizaje motor, habilidades básicas y complejas, competición, deportes, esfuerzo físico, calentamiento, cualidades físicas y condición física, así como los aspectos perceptivos y motrices del movimiento.
2. Actitudes y relación sociocultural. Expresión y comunicación (con 10 categorías). Constituye una dimensión relacionada con las actitudes y con las actividades expresivas y comunicativas. Aquí se encuadraron categorías relacionadas con el juego limpio, la igualdad y otros valores educativos, colaboración-cooperación, relación social, trabajo en equipo, desarrollo y enriquecimiento personal, comunicación-creatividad-expresión, cultura-tradición, sociedad y entorno, y una categoría general de valores-actitudes y normas o educación en valores.
3. Salud (con 10 categorías). La actividad física y otros aspectos relacionados con la salud de los escolares se introdujeron en esta dimensión. Las categorías fueron: higiene, hábitos de práctica física, evitar conductas nocivas o lesiones, calidad de vida, bienestar físico y mental, imagen y estética corporal, tensión-relajación y respiración, alimentación-hidratación, cuidado del cuerpo y hábitos posturales, así como una categoría genérica sobre salud.
4. Autonomía corporal y de la actividad física (con 10 categorías). En esta dimensión se incluyeron aspectos relacionados con hacer al alumno más autónomo en su propia concepción del cuerpo y en su motricidad, así como en su análisis del entorno. Las categorías fueron relación con el entorno y aplicación a la vida, autosuperación-autoexigencia, responsabilidad y perseverancia, autoconfianza-autoorganización, resolución de problemas, toma decisiones, desarrollo cognitivo, dosificación e intensidad del esfuerzo, descubrir-elegir, comprensión y exploración del cuerpo, lateralidad, control y ajuste, así como una categoría general de autonomía.
5. Recreación, motivación y juego (con tres categorías). Esta dimensión trataba de agrupar todas las referencias que se realizan en los Reales Decretos de Enseñanzas Mínimas sobre el aspecto lúdico de la actividad física. Las categorías fueron actividad física en tiempo libre, recreación-motivación, enfoque lúdico-vivencial y una categoría general sobre juegos (autóctonos, tradicionales, libres y reglados).

Antes de dejar como definitivo el sistema de categorías, se aclararon algunas cuestiones conceptuales, como por ejemplo las referencias relacionadas con el calentamiento, que podrían encuadrarse en la dimensión de aprendizaje motor y condición física o en la de salud, pero siempre optamos por incluirlas según la orientación mayoritaria en que aparecían en el texto. Cada unidad de análisis o fragmento de texto al que se asigna una categoría, estaba en función del tema al que se refería el documento legislativo. Así, en ocasiones, sólo con la aparición de un término (e.g., salud, calentamiento, habilidades básicas) era suficiente para incluir el código correspondiente, mientras que en otras ocasiones era la frase completa la que nos proporcionaba el contenido de dicha categoría (Marcelo, 1992).

Los textos analizados comparativamente fueron: a) Anexo II referente al área curricular de Educación Física del Real Decreto 1006/1991 de Enseñanzas Mínimas de Educación Primaria de la LOGSE, frente a b) Anexo II referente al área curricular de Educación Física del Real Decreto 1513/2006 de Enseñanzas Mínimas de Educación Primaria de la LOE.

3. Resultados

A continuación se presentan, en varias tablas, las frecuencias de las categorías temáticas y sus dimensiones. Presentamos estos resultados en números absolutos de las apariciones en los textos y en porcentajes relativos al número total de categorías de ese texto en concreto. Entendemos que ambos datos tienen valor interpretativo en nuestro estudio. Por un lado, el número absoluto demuestra indirectamente la preocupación administrativa de esas categorías en las leyes analizadas. Las reformas deben mostrar las nuevas tendencias y los avances científicos en los textos de sus leyes educativas, reflejándose en las enseñanzas mínimas de cada área curricular. Por otro, los porcentajes nos indican comparativamente con otras categorías la importancia que cada una tiene respecto al resto. Además, nos permiten comparar el peso específico de una categoría determinada entre ambas leyes sin tener en cuenta la longitud de cada texto (y su consecuente diferencia en el número absoluto de categorías, como es el caso de estas dos leyes).

La tabla 1 muestra los datos de las dimensiones que engloban a las 42 categorías, dándonos una idea general de los textos globales de cada ley. En las tablas 2, 4, 5 y 6, se presentan un análisis detallado de cada dimensión y las categorías que la componen, especificándose además en qué partes del documento se encuentran las frecuencias de dichas categorías (introducción, objetivos, contenidos, orientaciones metodológicas y criterios de evaluación). Finalmente, en la tabla 7 se muestra un análisis comparativo de los objetivos generales de la EF en la etapa de Primaria y de los Criterios de Evaluación en ambas leyes, ya que suponen dos apartados comparables en ambas leyes y de gran importancia en la orientación de la práctica docente.

Tabla 1. *Resultados generales de frecuencia de temas en los centros de interés de EF en Primaria detectados en ambos documentos. Frecuencia absoluta y porcentaje*

GRUPOS O CENTROS DE INTERÉS	LOGSE Frecuencia	/ %	LOE Frecuencia	/ %
Habilidades y destrezas-condición física-desarrollo motor	140	29,79	211	26,34
Actitudes y relación sociocultural. expresión-comunicación	134	28,51	242	30,21
Salud	72	15,32	95	11,86
Autonomía corporal y de la actividad física	83	17,66	183	22,85
Recreación, motivación y juego	41	8,72	70	8,74
TOTALES	470	100	801	100

En la tabla 1 vemos cómo la frecuencia de categorías en ambas leyes para la EF de Primaria es similar. Las principales diferencias se detectan en dos ámbitos. En primer lugar, mientras que en la LOGSE los porcentajes de los dos primeros grupos son muy similares y además los más importantes, con un 29,79% y 28,51%, en la LOE se distancian. En el actual Real Decreto de Enseñanzas Mínimas de EF para Primaria, parece dar más importancia a las 'Actitudes y relación sociocultural. Expresión y comunicación' distanciándose casi cuatro puntos sobre el grupo de 'Habilidades y destrezas. Condición física y desarrollo motor' que queda en segundo lugar con 26,34%. Por otra parte, el grupo de 'Autonomía corporal y de la actividad física' que tenía un porcentaje de 17,66% en la LOGSE, incrementa su frecuencia en algo más de cinco puntos, pasando a 22,85% en la LOE. También es destacable la disminución de frecuencia relativa ocurrida en el grupo de salud, donde del 16,28% de la LOGSE pasa al 11,73% en la LOE. Sin embargo, como veremos en el análisis de la tabla 7, estos resultados pueden profundizarse y la interpretación global de la importancia de la salud en la LOE se refleja en los sub-apartados.

Tabla 2. *Grupo 1 de habilidades y destrezas. Condición física y desarrollo motor.*

Dimensión 1: **Habilidades y destrezas-condición física-desarrollo motor**	INTRODUCCIÓN LOGSE / LOE		OBJETIVOS LOGSE/LOE		CONTENIDOS LOGSE/LOE		ORIENT. METODOL. LOGSE	CRITERIOS DE EVALUACIÓN LOGSE / LOE		TOTALES -Frecuencia absoluta/Porcentaje relativo LOGSE/LOE	
CAP-M-GRALES	15	14	2	3	18	16	9	6	18	50/35,71	51/24,17
HAB-BÁSICAS	3	-	1	-	11	2	3	8	44	26/18,57	46/21,80
HAB-ESPECÍF.	1	1	-	-	3	5	1	-	12	5/3,57	18/8,53
COMPETICIÓN	1	1	1	-	-	5	1	-	8	3/2,14	14/6,63
DEPORTE	3	8	1	2	1	9	2	-	5	7/5	24/11,37
ESFUERZO FÍSICO	-	1	1	1	-	2	2	1	4	4/2,86	7/3,32
CALENTAMIENTO	-	-	-	-	2	2	2	-	-	4/2,86	2/0,95
COND-FÍSICA	1	2	1	1	6	4	-	5	6	13/9,29	13/6,16
TAREAS-MOT	5	-	1	-	21	14	1	-	21	28/20	36/17,06
TOTALES	29	27	8	7	62	59	21	20	118	**140**	**211**

NOTA: CAP-M-GRALES (capacidades motrices generales), incluye capacidades de equilibrio estático y dinámico, práctica motriz, coordinación psicológica física y motriz, destrezas y aptitudes, desarrollo y aprendizaje motor, patrones motores, y uso eficaz del cuerpo; HAB-BÁSICAS incluye todas las habilidades básicas (e.g., saltos, giros, manejo objetos, lanzamientos, recepciones, bote); HAB-ESPECÍF. incluye a las habilidades complejas, específicas (puntería, interceptación, etc.) y combinación de ellas; COMPETICIÓN incluye referencias a las actividades de competición y oposición; DEPORTE incluye referencias al deporte adaptado y a las actividades deportivas en general; COND-FÍSICA se refiere a las cualidades físicas, condición física e intensidad movimiento; TAREAS-MOT (tareas motrices) se refiere a aspectos perceptivos y motrices del movimiento, funcionalidad, percepción espacio temporal (sentido, orientación, planos, simetría, volúmenes, frecuencia) y trayectorias.

El número absoluto de frecuencia de aparición de categorías en esta tabla 2 es superior en la LOE en casi todos los ámbitos. Los principales cambios surgen en los porcentajes de CAP-M-GENERALES, con casi un 12% menos en la LOE, el descenso de algo más del 3% en CON-FÍSICA, y los incrementos de las dos HAB, COMPETICIÓN y DEPORTE. El peso específico de esta dimensión en general dentro de ambos documentos es muy parecido como se mostró en la tabla 1.

En los criterios de evaluación aparecen HAB-ESPECÍF., COMPETICIÓN, DEPORTE y TAREAS-MOT, que no lo hacían en la LOGSE.

Tabla 3. *Grupo 2 de actitudes y relación sociocultural, expresión y comunicación.*

Dimensión 2: **Actitudes y relación sociocultural, expresión y comunicación**	INTRODUCC. LOGSE / LOE		OBJETIVOS LOGSE/LOE		CONTENIDOS LOGSE / LOE		ORIENT. METOD. LOGSE	CRITERIOS EVALUACIÓN LOGSE / LOE		TOTALES -Frecuencia/ Porcentaje relativo LOGSE/LOE	
ANTIVIOLENCIA	-	1	1	-	1	2	-	-	1	2/1,49	4/1,65
IGUALDAD	2	9	1	3	6	7	3	5	4	17/12,69	23/9,50
COLABORAC.	1	4	-	1	1	4	1	1	7	4/2,98	16/6,61
INTER-SOCIAL	11	11	2	1	7	4	4	2	10	26/19,40	27/11,16
T-EQUIPO	1	1	-	1	4	-	-	1	8	6/4,47	10/4,13
INTEGRAL	2	1	-	-	-	-	-	-	-	2/1,49	1/0,41
COMUNICACIÓN	8	16	3	3	16	44	3	13	34	44/32,84	97/40,08
CULTURA	1	8	-	1	4	2	2	-	1	7/5,22	12/4,96
SOCIEDAD	1	2	-	-	1	2	-	-	-	2/1,49	4/1,65
VAL-ACT-NOR	6	9	2	3	9	23	2	6	14	25/18,65	49/20,25
TOTALES	33	62	9	13	49	88	15	28	79	**134**	**242**

NOTA: ANTIVIOLENCIA incluye referencias a la no violencia en el deporte y la actividad física, deportividad y juego limpio; IGUALDAD incluye referencias a la no discriminación por ninguna razón, solidaridad, respeto y tolerancia a los demás; COLABORAC. incluye actividades de colaboración y cooperación; INTER-SOCIAL incluye a las relaciones sociales, interacción y adaptación social, habilidades sociales, y roles; T-EQUIPO se refiere al trabajo en equipo; INTEGRAL incluye el enriquecimiento y desarrollo personal e integral del individuo; COMUNICACIÓN incluye todo lo referente a creatividad y expresión (danza, mimo, representaciones, teatro, desinhibición, bailes y coreografías, capacidad expresiva); CULTURA incluye referencias a la tradición y a la cultura propias del entorno; SOCIEDAD incluye referencias a la sociedad general, y al entorno social cercano; VAL-ACT-NOR (valores, actitudes y normas) hace referencia a la educación en valores (tomar conciencia, valoración, disposición a…).

Aparte del incremento del número de veces que aparecen estas categorías en la LOE (108 veces más), los porcentajes en esta dimensión son similares. Se ha producido un incremento porcentual en la COMUNICACIÓN, COLABORAC., mientras que ha disminuido fundamentalmente en INTER-SOCIAL.

En los objetivos de la LOE se hace más hincapié en la IGUALDAD, mientras que en los criterios de evaluación, se incrementan INTER-SOCIAL, COMUNICACIÓN y ACT-VAL-NOR de manera importante.

Tabla 4. *Grupo 3 de categorías relacionadas con la salud.*

Dimensión 3: **Salud**	INTRODUCC.		OBJETIVOS		CONTENIDOS		ORIENT. METODOL.	CRITERIOS EVALUACIÓN		TOTALES Frecuencia/Porcentaje relativo	
	LOGSE	LOE	LOGSE	LOE	LOGSE	LOE	LOGSE	LOGSE	LOE	LOGSE	LOE
HIGIENE	2	-	1	1	5	3	2	-	2	10/13,89	6/6,32
HÁBITO-AF	1	6	1	-	2	-	-	1	2	5/6,94	8/8,42
SEGURIDAD	-	-	-	-	9	5	5	-	6	14/19,44	11/11,58
C-VIDA	2	1	-	-	2	-	-	-	-	4/5,56	1/1,05
BIENESTAR	-	5	-	1	1	2	-	-	1	1/1,39	9/9,47
IMAGEN	1	1	-	-	1	4	-	-	-	2/2,78	6/6,32
RELAJACIÓN	-	-	-	-	6	8	2	-	5	8/11,12	13/13,68
ALIMENTA.	1	-	1	1	2	3	2	-	3	6/8,33	7/7,37
CUIDADO	2	-	1	1	3	-	-	-	3	6/8,33	4/4,21
HÁBITO-SALUD	3	8	1	2	8	14	1	3	7	15/20,83	31/32,63
TOTALES	12	21	5	6	39	39	12	4	29	**72**	**95**

NOTA: HIGIENE se refiere al aseo, la higiene corporal y el vestuario; HÁBITO-AF se refiere a establecer hábitos de práctica física, y evitar el sedentarismo; SEGURIDAD hace referencia a evitar conductas nocivas, normas de seguridad, y evitar lesiones en la práctica de actividad física; C-VIDA se refiere a la calidad de vida; BIENESTAR hace referencia al estado mental y físico de bienestar de los alumnos; IMAGEN se refiere a la estética y la imagen corporal de los alumnos; RELAJACIÓN se refiere a los métodos de tensión y relajación muscular, así como a la respiración; ALIMENTA (alimentación) se refiere a los contenidos de hidratación y alimentación; CUIDADO se refiere a cuidado del cuerpo y hábitos posturales correctos; HÁBITO-SALUD hace referencia a los hábitos saludables y referencias a la salud en general.

Las referencias de HÁBITO-SALUD han supuesto el principal incremento de esta dimensión, con un importante casi 12% más en la LOE. La IMAGEN, así como el BIENESTAR, también incrementaron igualmente en la LOE. La HIGIENE, SEGURIDAD y C-VIDA aparecen en menor porcentaje. En los criterios de evaluación de la LOE esta dimensión sí ha tenido un gran incremento en casi todas sus categorías (de 4 a 29 referencias). En la tabla 7 se muestra un análisis complementario que confirma el incremento de la importancia de la salud en la LOE en estos criterios de evaluación.

Tabla 5. *Desglose de frecuencia categorías del centro de interés de autonomía corporal y de la actividad física*

Dimensión 4: **Autonomía corporal y de la actividad física**	INTRODUCC.		OBJETIVOS		CONTENIDOS		ORIENT. METODOL.	CRITERIOS EVALUACIÓN		TOTALES -Frecuencia -Porcentaje relativo	
	LOGSE	LOE	LOGSE	LOE	LOGSE	LOE	LOGSE	LOGSE	LOE	LOGSE	LOE
VIDA	4	6	-	1	3	5	6	3	27	16/19,28	39/21,31
AUTOSUPERA	-	2	1	3	1	1	-	-	4	2/2,41	10/5,46
AUTOCONFIA	-	3	-	-	-	-	-	-	2	-	5/2,73
CRÍTICA	5	8	3	5	6	2	2	2	42	18/21,69	57/31,15

DOSIFICA	-	-	1	1	2	2	3	1	1	7/8,43	4/2,19
DESCUBRIR	-	-	-	-	-	-	-	-	7	-	7/3,82
EXPLORA	9	11	4	1	20	28	-	4	8	37/44,58	48/26,23
AUTONOMÍA	-	5	-	1	2	1	-	1	6	3/3,61	13/7,10
TOTALES	18	35	9	12	34	39	11	11	97	**83**	**183**

NOTA: VIDA referencias a la aplicación de los aprendizajes al entorno (urbano y natural) y a su aplicación a la vida; AUTOSUPERA incluye referencias a la superación personal, autoexigencia, responsabilidad y perseverancia; AUTOCONFIA se refiere a la autoconfianza y a la autoorganización de tareas del alumnado; CRÍTICA se refiere a la actitud crítica, resolución de problemas, toma de decisiones y al desarrollo cognitivo y psíquico general; DOSIFICA hace referencia a las regulaciones en intensidad y dosificación del esfuerzo; DESCUBRIR se refiere a elaborar, elegir y descubrir por sí mismo; EXPLORA hace referencia a la comprensión, identidad y exploración del cuerpo (global y segmentaria), a la relación consigo mismo, posibilidades sensoriales, lateralidad y, control y ajuste corporal; AUTONOMÍA incluye autonomía general y la toma de iniciativas.

En la tabla 5 se muestra la dimensión que busca la autonomía del alumnado de Primaria, tanto en lo corporal como en la regulación de su actividad física. El cambio ha sido sustancial en el número de veces que se nombran sus categorías, con un incremento de 100 ocasiones. Aparecen incluso categorías nuevas como AUTOCONFIA y DESCUBRIR. Los mayores incrementos de porcentaje de aparición relativa se dan en CRÍTICA y AUTONOMÍA. Las categorías de EXPLORA y DOSIFICA disminuyen, sin embargo siguen apareciendo en los objetivos y en los criterios de evaluación de la LOE.

En los objetivos de la LOE aparecen reflejadas la VIDA y AUTONOMÍA respecto a la LOGSE. Sin embargo, disminuye su presencia en estos objetivos generales de la materia de EF la categoría de EXPLORA.

En los criterios de evaluación han habido apariciones nuevas de categorías en esta dimensión (la más numerosa ha sido DESCUBRIR con siete apariciones) y se han incrementado notablemente la categoría de CRÍTICA y VIDA. Además, como se observa en la tabla 7, esta dimensión de autonomía tiene mayor presencia porcentual relativa (más del 10%) en la LOE respecto a la LOGSE en los criterios de evaluación de esta etapa.

Tabla 6. *Frecuencia de categorías en el grupo de recreación, motivación y juego*

Dimensión 5: **Recreación, motivación y juego**	INTRODUCC.		OBJETIVOS		CONTENIDOS		ORIENT. METODOL.	CRITERIOS EVALUACIÓN		TOTALES -Frecuencia -Porcentaje relativo	
	LOGSE	LOE	LOGSE	LOE	LOGSE	LOE	LOGSE	LOGSE	LOE	LOGSE	LOE
T. LIBRE	4	2	1	1	1	2	1	-	-	7/17,07	5/7,14
RECREACIÓN	7	3	2	2	7	5	4	-	2	20/48,78	12/17,14
JUEGO	2	7	1	-	8	25	3	-	21	14/34,15	53/75,71
TOTALES	13	12	4	3	16	32	8	-	23	**41**	**70**

NOTA: T. LIBRE se refiere a la actividad física realizada en su tiempo libre y al ocio; RECREACIÓN se refiere a todo lo relativo a la motivación, el disfrute, y el carácter recreativo y lúdico de las vivencias en la actividad física; JUEGO incluye referencias a juegos autóctonos, tradicionales, libres y reglados. En esta dimensión relacionada con el aspecto lúdico y motivacional del juego casi se han duplicado las referencias absolutas, fundamentalmente debido al incremento de la categoría JUEGO, mientras que las otras han disminuido. Sin embargo, los porcentajes relativos en cada ley de esta dimensión han sido prácticamente los mismos (ver tabla 1). Finalmente, queremos comentar que aunque en la LOE no existe un apartado de orientaciones metodológicas como en la LOGSE, sí se mencionan fundamentalmente el uso de materiales multifuncionales; la adecuación al alumno y la adaptación del trabajo; la globalidad de esta etapa y su unión con otras áreas; así como la utilización del juego.

Para una correcta interpretación de las intenciones de ambas leyes es aconsejable realizar una valoración ponderada de los elementos del currículo analizados en ambos documentos. Desde esta perspectiva, los objetivos del área de EF en esta etapa y los criterios de evaluación, ya que marcan el producto a conseguir con los alumnos de esta etapa, son de mayor importancia que por ejemplo el apartado introductorio o incluso que los mismos contenidos, y por tanto determinantes en la manera de actuar de los docentes. Por esta razón presentamos en la tabla 7 la aparición temática comparativa entre las dos leyes en estos dos elementos.

Tabla 7. *Comparación de la frecuencia y porcentaje relativo de aparición de temas en los objetivos generales de la EF de primaria y criterios de evaluación en las dos leyes*

	Objetivos Generales de EF		Criterios de Evaluación				
	LOGSE	LOE	LOGSE	LOE			
				1º ciclo	2º ciclo	3º ciclo	Total LOE
Nº total	8 (100%)	8 (100%)	15 (100%)	8	8	8	24 (100%)
Dimensión 1	5 / 62,5% (3,4,5,6,7)	5 / 62,5% (1,3,4,5,8)	9 /60% (1,3,4,5,6,7,8,9,15)	4 (1,2,3,5)	6 (1,2,3,5,6,8)	5 (1,2,4,5,6)	11 / 45,83%
Dimensión 2	5 / 62,5% (1,2,6,7,8)	5 / 62,5% (1,4,6,7,8)	5 / 33,33% (2,10,13,14,15)	3 (5,6,7)	4 (4,5,6,7)	6 (2,3,4,5,7,8)	13 / 54,17%
Dimensión 3	1 / 12,5% (2)	2 / 25% (2,6)	1 / 6,67% (12)	2 (4,8)	2 (3,8)	1 (8)	5 / 20,83%
Dimensión 4	6 / 75% (1,2,3,4,5,7)	7 / 87,5% (1,2,3,4,5,7,8)	4 / 27,33% (1,3,7,11)	2 (1,2)	2 (1,4)	5 (1,2,3,4,5)	9 / 37,5%
Dimensión 5	2 / 25% (1,6)	2 / 25% (1,8)	5 / 33,33% (6,8,13,14,15)	2 (3,5)	4 (2,4,5,8)	3 (2,3,4)	9 / 37,5%

NOTA: Se presenta la frecuencia por cada dimensión y entre paréntesis el número del objetivo o del criterio de evaluación concreto en el que aparece.

El número de objetivos generales de la EF en primaria no ha variado en ambas leyes (8). El número y frecuencia en las dimensiones temáticas dentro de los objetivos son similares. En la LOE aparece una mención más a la salud (dimensión 3) y una más a la autonomía (dimensión 4), que incrementan su porcentaje relativo de 12,5 a 25% y de 75% a 87,5%, respectivamente.

Como se observa, y a pesar de que ambos elementos de los Reales Decretos son comparables, en la LOE aparecen divididos los criterios de evaluación por ciclos (al igual que los contenidos), y por tanto se multiplica la aparición de temas y el número total de criterios, pasando de 15 en la LOGSE a 24 en la LOE (8 en cada ciclo). El mayor incremento porcentual relativo corresponde a las dimensiones 2 y 3, pasando de 33,33 y 6,67% en la LOGSE a 54,17 y 20,83% en la LOE, respectivamente. A pesar del número mayor de criterios de evaluación en la LOE, la dimensión 1 disminuye su porcentaje relativo de aparición en ellos, pasando del 60% en la LOGSE al 45,83% en la LOE.

4. Discusión

Como en todas las investigaciones de este tipo, el nivel de abstracción de las categorías representativas de los conceptos que aparecen en los documentos analizados supone un proceso subjetivo, que depende de los investigadores. La inducción de estas categorías temáticas siguen un proceso sistemático que minimiza el error, y está basada en 'no interpretar' el texto, sino en identificar los temas explícitamente señalados en él (Viciana y Sánchez, 2002). Sin embargo, siempre existe un componente cognitivo y conceptual previo de los investigadores, que condiciona la definición de estas categorías temáticas. Por ello, más que una limitación del estudio, lo destacamos como un componente identificativo del mismo del que debemos ser conscientes.

También cabe destacar que la comparación absoluta de las frecuencias de los temas tratados en ambas leyes están relacionadas con la extensión de cada documento, así como con los apartados de cada una, que no son exactamente iguales (véase el apartado de metodología en LOGSE que no aparece en LOE). Esto provoca que la comparación no pueda ser tenida en cuenta como exacta, sino como referencia. Los porcentajes de aparición de los temas reflejados en las tablas de resultados palían brevemente este desajuste entre ambas leyes, sin ser la solución definitiva para la comparación absoluta.

Además de otros cambios más superficiales y fácilmente detectables como la incorporación de las competencias básicas o el tratamiento de los contenidos diferenciándolos por ciclos (Herrador, Huertas y Lara, 2009), hemos realizado este análisis de contenido de los dos Reales Decretos con el fin de identificar otros cambios más encubiertos en el propio texto. Deducimos por tanto, que los temas tratados en ambos documentos y la frecuencia de aparición de cada uno de ellos reflejan la importancia de dichos temas para la sociedad española de ambos momentos.

Viciana, Salinas y Cocca (2007) realizaron este mismo análisis con los Reales Decretos relativos a la etapa de Secundaria. Los resultados mostraron que el enfoque del currículo de la LOE estaba más cerca del producto que el de la LOGSE, donde la EF se aproximaba más al proceso y el currículo era más indeterminado y abierto. Esta indeterminación en el currículo de primer nivel de EF, tanto en Primaria como en Secundaria, había provocado una cierta inestabilidad y una falta de coherencia en la progresión horizontal (organización o secuencia en complejidad del área) de los contenidos de las etapas (Viciana, Zabala, Sánchez y Lozano, 2004), cuestión que se soluciona ahora en la LOE marcando esta progresión en los ciclos de Primaria (Real Decreto 1513/2006) y en los cursos de Secundaria (Real Decreto 1631/2006). Además, el primer nivel de concreción de la LOE en Secundaria se enfoca aún más hacia los procedimientos en lo referente a "Habilidades y destrezas. Condición física y desarrollo motor", en detrimento de la dimensión de "Actitudes y relación socio-cultural" que tuvo el predominio en la LOGSE (Viciana et al., 2007). Precisamente ocurre lo contrario que en este análisis llevado a cabo en el currículo de Primaria, donde se concede más importancia a esta dimensión de "Actitudes y relaciones socio-culturales" frente al aprendizaje de "Habilidades motrices y desarrollo motor" con edades más tempranas. Quizá sea lógico que en esta etapa, las relaciones sociales y la actitud sean más importantes que el aprendizaje motor, mientras que en la siguiente el niño-adolescente ya comienza a desarrollar más su condición física y las habilidades específicas deportivas. En la redacción del currículo estándar nacional de otros países desarrollados como Estados Unidos, también se hace hincapié sobre la progresión del acondicionamiento físico desde Primaria, aunque como resaltan algunos autores (Newell, 2011), el problema está en el paso a la práctica, o sea, saber cuántos profesores realmente lo están llevando a cabo y de qué manera.

Sin embargo, el hecho que la LOE muestre con una considerable mayor frecuencia las categorías de la dimensión de "Habilidades y destrezas. Condición Física y desarrollo motor" en los criterios de evaluación de Primaria, confirma que la actual Ley educativa otorga gran importancia a este aprendizaje en nuestro Área de EF. Sabemos que la LOGSE, como reacción a la LODE anterior, mostró cierta reticencia a los logros de aprendizaje motor o de condición física (normalmente acompañada de una evaluación normativa), y apostó por las vivencias, la recreación y la participación del alumnado en las clases de EF (Viciana, 2002). Sin embargo se ha vuelto a replantear, a juicio de los resultados obtenidos en este análisis, la verdadera esencia del área retomando la importancia de esta parte de la EF. Esta preocupación es actualmente extensible a otros países que reflejan el interés global por este tipo de aprendizaje (Newell, 2011). Los profesores de EF han ido reclamando progresivamente en estos años atrás, que los alumnos deben aprender en sus clases, que el aprendizaje deportivo y de control del esfuerzo físico en la vida diaria de los niños es importante para la salud física y psicológica, que conlleva a una mayor calidad de vida en el futuro. Por el contrario, asistir a clase de EF para "liberarse" del esfuerzo cognitivo que suponen las asignaturas más teóricas ha perdido peso y sentido (aunque esta función catártica de la actividad física se realiza por sí misma), y ya las experiencias gratificantes debemos coordinarlas con los aprendizajes motores y cognitivos que nos permitirán conseguir mayores logros.

La formación de ciudadanos activos, críticos, responsables y sociales es un objetivo de la actual Reforma Educativa, evidenciado en la nueva materia de Educación para la Ciudadanía y en las competencias básicas a desarrollar en Primaria y Secundaria. En nuestro análisis del área de EF en Primaria, se ha reflejado en el cambio de la dimensión de "Actitudes y relaciones socio-culturales" mostrándose como la más frecuente en el Real Decreto de Enseñanzas Mínimas. En España se están realizando actualmente investigaciones en este ámbito que evidencian la importancia de las relaciones sociales en la EF (Pascual, Escartí, Llopis, Gutiérrez, Marín y Wright, 2011; Estrada, González-Mesa, Méndez-Giménez y Fernández-Río, 2011).

La salud, uno de los principales objetivos de la educación en general, también lo es para la EF de Primaria. Aunque se produjo un decremento de cinco puntos porcentuales en la frecuencia relativa de la LOGSE a la LOE, es cierto también que aumentaron considerablemente de manera absoluta en casi todas las categorías de esta dimensión en los criterios de evaluación. Creemos que no es necesario recordar la importancia de esta dimensión para esta etapa puesto que los profesores lo tienen muy presente en sus innovaciones docentes diarias (Mayorga, Viciana, Cocca y Miranda, 2010). Además, esta etapa es especialmente importante para mantener los niveles de actividad física, fundamentalmente cuidando el ambiente que rodea a los alumnos (Knowles, Niven y Faukner, 2011), con el fin de que no decaiga en la transición a Secundaria como se ha demostrado en investigaciones cuantitativas (Cocca, Viciana, Salinas, Salazar, Medina y Miranda, 2009).

El incremento sustancial de la dimensión de "Autonomía corporal y de actividad física" en la LOE refleja la importancia de este tema para la sociedad española actual en la etapa de Primaria, llegando a su máxima expresión en el último curso de Secundaria (Viciana et al., 2007). La competencia básica de autonomía e iniciativa personal proveniente de la

Educación por Competencias (McClelland, 1973), uno de los pilares básicos de esta Reforma Educativa en la LOE, refleja también esta importancia de dotar de autonomía al alumnado en estas etapas. Otras investigaciones que analizan el currículo estándar nacional de sus países respectivos, coinciden en dichos objetivos hacia el alumnado de Primaria, resaltando que el problema es la cualificación del profesorado para orientar al alumno hacia el futuro en EF (Van Deventer y Van Niekerk, 2009).

La motivación es uno de los pilares de la enseñanza efectiva para el aprendizaje en cualquier etapa educativa, pero quizá aún más en Primaria (Chedzoy y Burden, 2009). Aunque la frecuencia de aparición entre LOGSE y LOE es prácticamente la misma, creemos que su frecuencia relativa respecto a las demás dimensiones es escasa, tratándose de la etapa de Primaria, donde la metodología recomendada debe basarse en juegos y en la motivación del alumnado.

En nuestra opinión, el verdadero problema de las Reformas educativas siempre será la implementación de los cambios propuestos. Se ha demostrado que muchos problemas de estas implementaciones provienen de: la interpretación de estos cambios, para lo cual hemos desarrollado este trabajo, aportando una visión más clara y objetiva de la evolución de la EF en Primaria de la LOGSE a la LOE; de la formación que reciben los profesores para realizarlos; del control y asesoría de la Administración en los centros educativos cuando aplican los cambios establecidos (Keating, Lambdin, Harrison y Dauenhauera, 2010); y de las funciones docentes que son susceptibles de mejora en los profesores de EF para aplicarlos (Morgan y Hansen, 2007). Medidas como las de dejar un porcentaje del tiempo del horario lectivo para que los profesores en sus centros puedan planificar (Blair y Capel, 2008) o hablar con sus compañeros de diferentes tipos de intervenciones, podrían servir para enriquecer la enseñanza y ayudar a interpretar e implementar los cambios propuestos por las Reformas Educativas generales.

Además, como conclusiones prácticas derivadas de este estudio podemos destacar las siguientes:

- Los maestros de EF podrían enfocar prioritariamente su enseñanza hacia las “Actitudes y relaciones socio-culturales”, ya que es la más relevante de esta etapa según la frecuencia de sus categorías. Lógicamente no olvidamos que las finalidades educativas de los centros también condicionarán esta prioridad en la práctica.
- El aprendizaje de “Habilidades motrices y desarrollo motor” tan característico de la EF es el siguiente núcleo en importancia. Usar estilos de enseñanza individualizadores como los grupos de nivel e intereses, así como crear un buen clima social y de aprendizaje serán fundamentales para conseguir estos objetivos. La individualización permitirá que los alumnos trabajen a su ritmo para conseguir los objetivos de aprendizaje motor más rápidamente y provocarán que el tiempo de compromiso motor en clase sea exitoso.
- La recreación y el juego serán los mejores vehículos para conseguir el objetivo anterior de aprender con motivación. Por ello, diseñar tareas con una meta alcanzable, usar el grupo para colaborar hacia un fin común, proponer competiciones colaborativas, facilitar elementos técnicos en su ejecución, facilitar las tácticas y estrategias de los pre-deportes con adaptaciones de los campos de juego y otros recursos, serán igualmente importantes estrategias de motivación para lograrlo.
- La dimensión de “Autonomía corporal y de la actividad física” es ahora más importante. Usar los conocimientos del aula para adaptarlos a la vida con aplicaciones

en forma de actividades complementarias sería un buen ejemplo de llevar a la práctica este resultado. Usar el entorno urbano y natural del niño para aplicar actividades físicas (en las plazas públicas, en la playa o en la montaña) con sus familias y controlando el esfuerzo, serían ejemplos viables y de acorde a esta conclusión.

- Otras conclusiones más particulares y derivadas de las categorías de cada dimensión podrían ser aplicables por los docentes, y por tanto dejamos a sus criterios dichas aplicaciones.

Esperamos que los resultados mostrados en este trabajo puedan aclarar al profesorado de EF los cambios más encubiertos que sugiere el primer nivel de concreción del currículo español en EF. Desde esta perspectiva, esperamos que el profesorado pueda aplicar de manera más fácil y acorde a lo sugerido por la Administración en esta etapa de Primaria, lo que la Ley educativa les está solicitando.

Referencias bibliográficas

Blair, R. y Capel, S. (2008). Intended or unintended? Issues arising from the implementation of the UK Government's 2003 Schools Workforce Remodelling Act. *Perspectives in Education*, *26*(2), 105-121.

Chedzoy, S. y Burden, R. (2009). Primary school children's reflections on Physical Education lessons: An attributional analysis and possible implications for teacher action. *Thinking Skill and Creativity*, *4*(3), 185-193.

Cocca, A., Viciana, J., Salinas, F., Salazar, C., Medina, R. T. y Miranda, Mª. T. (2009). Correlación entre actividad física y esquema corporal en jóvenes españoles de 8 a 23 años. *Revista Mexicana de Psicología*, *26*, 577-579.

Go Goetz, J. P. y LeCompte, M. D. (1984). *Ethnography and qualitative design in educational reseach*. New York, USA: Academic Press.

Herrador, M., Huertas, J. A. y Lara, C. (2009). LOE and LOGSE. A change of perspective in physical education. *The International Journal of Medicine and Science in Physical Education and Sport, 5*(1), 13-24.

Keating, X. D., Lambdin, D., Harrison, Jr. L. y Dauenhauera, B. (2010). Changes in K-12 physical education programs from 2001 to 2006. *Research Quarterly for Exercise and Sport*, *81*(2), 180-188.

Knowles, A. M., Niven, A. y Faukner, S. (2011). A qualitative examination of factors related to the decrease in physical activity behavior in adolescent girls during the transition from primary to secondary school. *Journal of Physical Activity and Health*, *8*(8), 1084-1091.

Learreta, B. (2003). *Nuevas enseñanzas mínimas para la EF de ESO: análisis comparativo entre el anterior y el nuevo currículo*. *Revista Española de Educación Física*, *374*, 42-47.

Ley Orgánica 1/1990 de 3 de Octubre, de Ordenación General del Sistema Educativo (LOGSE).

Ley Orgánica 2/2006 de 3 de Mayo, de Educación (LOE).

Ley Orgánica 8/1985, de 3 de julio, Reguladora del Derecho a la Educación (LODE).

Marcelo, C. (1992). *Desarrollo profesional e iniciación a la enseñanza*. Madrid: PPU.

Mayorga, D., Viciana, J., Cocca, A. y Miranda, Mª. T. (2010). Efecto del entrenamiento de la fuerza sobre la resistencia muscular abdominal en escolares pre-púberes. *Lecturas: Educación Física digital*, *15*(148). Disponible en http://www.efdeportes.com/efd148/resistencia-muscular-abdominal-en-pre-puberes.htm

McClelland, D. C. (1973). Testing for competence rather than for 'intelligence'. *American Psychologist*, *28*, 1-14.

Miles, M. y Huberman, A. (1984). *Qualitative data analysis.* London: SAGE Publications.

Muñoz, J. C. (2007). La Educación Física en la Ley Orgánica de Educación. *Lecturas: Educación Física y Deportes*, 105. Disponible en http://www.efdeportes.com/efd105/laeducacion-fisica-en-la-ley-organica-deeducacion.htm

Morgan, P. y Hansen, V. (2007). Recommendations to improve primary school physical education: Classroom teachers' perspective. *Journal of Educational Research*, *101*(2), 99-111.

Newell, K. M. (2011). Physical education of and through fitness and skill. *Quest*, *63*(1), 46-54.

Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Primaria Obligatoria.

Real Decreto 1006/1991, de 14 de junio, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Primaria Obligatoria.

Van Deventer, K. J. y Van Niekerk, E. (2009). Life orientation in the foundation phase (Grades R-3): A survey in selected Western Cape primary schools. *South African Journal for Research in Sport, Physical Education and Recreation*, *31*(2) 147-162.

Viciana, J. (2000). Principales tendencias innovadoras en la educación física actual. El avance del conocimiento curricular en educación física. *Lecturas: Educación Física y Deportes*, *19*. Disponible en http://www.efdeportes.com/efd19/innova.htm

Viciana, J. (2002). *Planificar en Educación Física*. Barcelona: Inde.

Viciana, J. y Sánchez, D. L. (2002). Procedimiento de inducción y aportación de un sistema múltiple de categorías para el análisis del discurso de entrenadores de deportes colectivos. *Lecturas: Educación Física y Deportes*, *53*. Disponible en http://www.efdeportes.com/efd53/discurs.htm

Viciana, J., Salinas, F. y Cocca, A. (2007). Análisis de contenido comparativo del primer nivel curricular de Educación Física en Secundaria. *Revista de currículum y formación del profesorado*, *11*(2). Disponible en http://www.ugr.es/~recfpro/rev112ART4.pdf

Viciana, J., Zabala, M., Sánchez, C. y Lozano, L. (2004). La evaluación de la Educación Física en la Ley de Calidad. Atajar la carencia de identidad de área. En C. Pablos y A. Rivera, *Jornades sobre l'evolució de l'activitat física il'esport*, cap. 13. Valencia: Servicio de Publicaciones de la Universidad de Valencia.

Agradecimientos

Agradecemos a Aliisa Hatten la revisión del inglés usado en el abstract. "Daniel Mayorga-Vega recibe una ayuda del programa de Formación del Profesorado Universitario (FPU) del Ministerio de Educación, Cultura y Deporte (AP2010-5905)"

Paper II

CUMPLIMIENTO DE LOS ESTÁNDARES CURRICULARES DE CONDICIÓN FÍSICA-SALUD EN EDUCACIÓN FÍSICA. ESTUDIO DE LA PLANIFICACIÓN EN LA FORMACIÓN INICIAL

[COMPLIANCE OF CURRICULUM STANDARDS OF HEALTH-RELATED PHYSICAL FITNESS IN PHYSICAL EDUCATION. STUDY OF THE PLANNING IN PRE-SERVICE TEACHERS]

Viciana, J., Mayorga-Vega, D., & Mompeán Campillo, M.

Cultura, Ciencia y Deporte

Pending acceptance of the revised version

Cumplimiento de los estándares curriculares de condición física-salud en Educación Física. Estudio de la planificación en la formación inicial

Compliance of curriculum standards of health-related physical fitness in Physical Education. Study of the planning in pre-service teachers

Resumen

Los objetivos del estudio fueron analizar la planificación del profesorado en formación inicial sobre condición física-salud, identificando errores y proponiendo recomendaciones para su mejora, así como crear un índice que indicara la calidad-efectividad al planificar la condición física-salud según los estándares nacionales en secundaria. La muestra se formó por 180 profesores (136 hombres y 34 mujeres) en formación inicial con edades entre 19 y 38 años (media 20,66 ± 2,94 años). Se crearon 31 categorías de análisis para categorizar diferentes parámetros de la planificación de la condición física-salud, y se crearon nuevas categorías a partir de aquellas. Los resultados indicaron que los profesores en formación inicial otorgan más importancia a la condición física-salud que a los demás contenidos, pero su planificación es deficiente (sesiones insuficientes y falta de refuerzos intermitentes) y por tanto inefectiva. Es necesario concienciar al profesorado de educación física en formación en una correcta planificación de la condición física-salud, implementándola y evaluándola en el practicum de la formación inicial.

Palabras clave: Profesores de educación física en formación inicial, educadores físicos, evaluación de programas, condición física, programación.

Abstract

The goals in this study were to analyze the planning of pre-service physical education teachers regarding the health-related physical fitness, identifying mistakes and proposing recommendations in order to improve the planning, as well as to make an index to determine the effectiveness-quality of a given planning in regard with health-related physical fitness, according to national secondary school standards. The sample was composed of 180 pre-service teachers (136 men and 34 women) whose age ranged from 19 to 38 years old (average age = 20.66 ± 2,94 years). In order to register and to categorize the different planning parameters of health-related physical fitness, 31 categories were created, and other categories were then created out of those 31 categories so as to complete the analysis. Results showed that pre-service teachers place great importance on health-related physical fitness, but their planning is deficient (they do not carry out enough sessions and lack intermittent reinforcements) and therefore is ineffective. It is recommended to be aware of educating pre-service teachers in a correct health-related physical fitness planning, implementing and evaluating it throughout the partnership of initial training.

Key words: Physical education pre-service teachers, physical educators, program evaluation, physical conditioning, physical performance, programming, programs.

1. Introducción

Uno de los retos de la sociedad del siglo XXI es incrementar la actividad física en la población española en edad escolar, debido a que conlleva enormes beneficios en los niños-jóvenes. Además el nivel de actividad física saludable establecido durante estas edades influirá a su vez en el estatus de salud y la calidad de vida durante la adultez (Ortega, Ruiz, Castillo, & Sjöström, 2008; Ruiz et al., 2009). Sin embargo, se ha demostrado que la actividad física decae durante las edades escolares, siendo este decremento especialmente relevante en la transición a la etapa de secundaria (Cocca, Liukkonen, Mayorga-Vega, & Viciana, 2014). Debido a esta contrastada tendencia negativa, la propia Organización Mundial de la Salud ha manifestado su preocupación social (Organización Mundial de la Salud, 2011) y sus recomendaciones de promoción de la salud se manifiestan actualmente en la mayor parte de las políticas de salud pública (Tremblay, 2014).

Asociada a la promoción y al hábito de actividad física saludable se encuentra la condición física, que es considerada un marcador de salud en los niños (Ortega et al., 2008). Además de otros beneficios psicológicos y sociales (Hallal, Victora, Azevedo, & Wells, 2006), se ha demostrado que los niños con un mayor nivel de condición física tienen igualmente mayores niveles de salud (Ortega et al., 2008). Además, la condición física ha demostrado ser un marcador de salud más potente incluso que otros más clásicos como por ejemplo la obesidad (Blair, 2009). Esto implicaría que independientemente del sobrepeso u obesidad, tener mayor nivel de condición física conllevaría igualmente un mayor nivel de salud.

Por ello, los estándares curriculares y las finalidades educativas de la mayor parte de los países desarrollados (National Association for Sport and Physical Education, 2004; Ministerio de Educación, Cultura y Deporte, 2007, 2014) contemplan este incremento y mantenimiento de condición física entre los escolares.

El papel de la educación es por tanto un objetivo fundamental para los gobiernos y sus sociedades, que consideran fundamentalmente al área curricular de Educación Física (EF) como una de sus principales estrategias para asegurar un nivel de condición física saludable entre la población escolar. En España, tanto el currículo de enseñanzas mínimas de secundaria de la Ley Orgánica de Educación (Ministerio de Educación, Cultura y Deporte, 2007), como el reciente currículo básico de la Ley Orgánica para la Mejora de la Calidad Educativa (Ministerio de Educación, Cultura y Deporte, 2014) han contemplado y mantenido estos importantes objetivos relacionados con la salud, reconociendo el papel del docente de EF como fundamental para la consecución de esta meta.

Sin embargo, planificar este incremento de la condición física-salud (CF-S) en el área de EF no es tarea fácil. El gran número de factores que influyen en ella (por ejemplo: solo dos días a la semana de práctica; vacaciones escolares que interrumpen las intervenciones docentes; la convivencia de muchos otros contenidos necesarios; o la heterogeneidad de los escolares que integran el grupo de clase) hace que planificarla eficazmente sea una tarea verdaderamente compleja. La necesidad de tratar un gran número de contenidos durante el año escolar da lugar a la creación de unidades didácticas (UD) extremadamente cortas, a veces sin sentido ni eficacia (Robles, Giménez, & Abad, 2010), provocando un desarrollo de la CF-S insuficiente (Pérez-Pastur, 2010). Además, las UD se diseñan como un número particular de lecciones basadas, en el mejor de los casos, en la propia experiencia del profesor y en el peor de ellos, en la mera intuición e incluso en la improvisación (Viciana & Zabala, 2004).

Hasta la actualidad, se han realizado en el contexto escolar diferentes programas que han obtenido éxito en el incremento de la CF-S, entiéndase el incremento de las cualidades de resistencia aeróbica, fuerza-resistencia y flexibilidad, tal y como señalan los criterios de evaluación de la legislación actual (Ministerio de Educación, Cultura y Deporte, 2007). Mayorga-Vega, Viciana, y Cocca (2013) lograron incrementar la resistencia aeróbica y fuerza-resistencia en la educación primaria, con un programa de 14 sesiones basado en ejercicios de alta intensidad aplicados en clase de EF. En esta investigación se mantuvieron altos niveles de tiempo de compromiso motor usando el circuito como sistema de organización, ya que había evidencias de ser el más efectivo (Viciana, Lozano, Cocca, & Mayorga-Vega, 2012). Y por último, para el incremento de la flexibilidad también se han realizado programas exitosos en EF de duración variable, desde ocho semanas hasta programas aplicados durante todo el curso escolar (Merino-Marban, Mayorga-Vega, Fernandez-Rodríguez, Vera-Estrada, & Viciana, 2015; Rodríguez, Santonja, López-Miñarro, Sáinz de Baranda, & Yuste, 2008). Estas evidencias empíricas hacen que sea real la posibilidad de incrementar la CF-S con programas de intervención relativamente cortos, cuestión que muchos docentes de EF creían imposible dados los problemas implícitos antes enumerados del contexto escolar (Salinas, 2011).

Sin embargo, las ganancias de CF-S que algunos docentes consiguen en EF deben ser mantenidas a lo largo del curso escolar, administrando nuevos estímulos cada cierto tiempo de desentrenamiento. Así se evitaría que los escolares volvieran a los valores basales, haciendo inútil el esfuerzo inicial que la incrementó. El modelo de aprendizaje exitoso de Viciana, Mayorga-Vega, y Cocca (2014) explica y desarrolla estos conceptos relacionados con el incremento y mantenimiento de la CF-S en el contexto escolar, a través de los denominados refuerzos intermitentes de Le Ny (1980). Incluso se han realizado estudios para determinar el tiempo en que el incremento de CF-S se pierde en los estudiantes (generalmente por las vacaciones escolares y por los periodos de impartición de otros contenidos del currículo que funcionan como desentrenamiento). Por ejemplo, Merino-Marban et al. (2015) hallaron que tras la aplicación de un programa de desarrollo, después de cinco semanas de desentrenamiento, la flexibilidad volvía a los valores iniciales en los estudiantes de primaria.

En la planificación de la EF hay también otros elementos importantes a considerar en el desarrollo de una actividad física regular y que influyen en la condición física. La utilización del entorno físico del centro, por ejemplo, es fundamental para que los estudiantes puedan usarlo en su tiempo extraescolar con conocimiento suficiente y autonomía; o la aplicación de nuevas tecnologías al alcance de los escolares y que sirven como ayuda metodológica para implementar correctos programas de entrenamiento (por ejemplo, aplicaciones de smartphones, pulsómetros o podómetros); e incluso la asociación de los contenidos curriculares de CF-S con las competencias.

Así pues, la necesidad de planificar detenidamente la CF-S parece evidente, siguiendo cuidadosas recomendaciones para garantizar cierto éxito (por ejemplo, duración de los programas de intervención, repetición del estímulo para evitar la pérdida o seguir las recomendaciones de la etapa y curso según el primer nivel de concreción del currículo). Desafortunadamente hasta la actualidad, solo existen estudios cualitativos sobre la planificación en EF centrados en las decisiones de planificación y describiendo el contexto de muestras pequeñas (casos), que indudablemente son útiles para mejorar la labor educativa de los profesores en formación inicial estudiados (Placek, 1984; Emmer, 1986). Sin embargo, dichos estudios no ofrecen datos generales que nos aporten una idea de si dicha planificación se está realizando correctamente. Por ello, la planificación de la CF-S en EF debe ser objeto de estudio y evaluación. Es necesario realizar un diagnóstico de los programas que los profesores en formación inicial realizan para sus clases con el fin de comprobar su efectividad potencial y el cumplimiento de los estándares legislativos actuales. Consecuentemente, podremos actuar en la formación de docentes promocionando la salud de los escolares y una mejor calidad de vida de la población española para un futuro próximo.

Los objetivos de este estudio fueron dos: (a) analizar las tendencias de los profesores de EF en formación inicial al planificar la CF-S, determinando errores y proponiendo recomendaciones que nos ayuden en el futuro a intervenir desde la propia universidad; y por otro lado, (b) crear un índice que nos indique la calidad-efectividad de una planificación en la programación de la CF-S.

2. Método

Se utilizó un diseño descriptivo transversal, basado en el análisis de la CF-S de las planificaciones escritas de profesores de EF de secundaria en formación inicial, así como algunos aspectos generales de la misma. La principal dificultad de este tipo de estudios radica en que los documentos son privados y por tanto, a veces, causantes de recelo por parte del profesorado a ser estudiados, pero con la ventaja de estar manejando documentos primarios y con el consecuente valor añadido a la investigación (Quintana, 2006).

2.1. Muestra

Intervinieron 211 profesores en formación inicial de la Universidad de Granada, de los cuales 180 conformaron la muestra definitiva (se eliminaron 31 planificaciones por contener defectos importantes de contenido que impedían su análisis para esta investigación). Del total, 146 fueron hombres (81,1%) y 34 mujeres (18,9%), con una media de edad de 20,66 (*DT* = 2,94). Los participantes cursaron la asignatura de Fundamentos de la EF de segundo curso del grado en Ciencias del Deporte durante los cursos 2012-13-14. Todos los participantes firmaron un consentimiento una vez informados del estudio.

2.2. Instrumento

Las categorías de análisis de las planificaciones constituyeron el instrumento de medida. Dichas categorías fueron determinadas por dos expertos en planificación y un profesor de EF de Secundaria. Para garantizar la validez del sistema de categorías y sus códigos de registro para esta investigación se realizaron seis sesiones de entrenamiento (análisis, definición y discusión). En ellas se codificaron cinco planificaciones escogidas al azar de entre la muestra del estudio donde se definieron-redefinieron todas las categorías de manera excluyente, determinando diferentes valores para cada una de ellas según cada caso. Se analizaron las coincidencias, discrepancias e inconsistencias en las categorías, se modificaron las definiciones y/o codificaciones con acuerdos entre los expertos hasta llegar a una coincidencia mayor del 80% en las tres últimas sesiones de entrenamiento, asegurando la consistencia del sistema (Anguera, 1988). Igualmente para asegurar la estabilidad de la codificación, se realizaron dos medidas de entrenamiento intra-codificador superando una coincidencia del 95%.

Las categorías de análisis fueron las siguientes:

- Número total de UD, y de UD y sesiones de CF-S.
- Trimestre donde la UD de CF-S es programada.
- Número de sesiones de fuerza-resistencia, resistencia aeróbica o flexibilidad y de refuerzos intermitentes.
- Número de objetivos dedicados a conceptos, procedimientos y actitudes en la UD de CF-S.
- Número de objetivos formulados erróneamente como contenidos y viceversa.
- Número de objetivos correctos/incorrectos según la concreción. Concreción incorrecta supone que se formula el objetivo a niveles más generales, tal y como aparecen en el Real Decreto de enseñanzas mínimas del primer nivel curricular.
- Estrategias de concreción empleadas en el objetivo para especificar su concreción.

- Criterios de evaluación del Real Decreto para tercero de Educación Secundaria Obligatoria (ESO).
- Número de contenidos adecuados y no adecuados a tercero de ESO en la UD y en las sesiones. Referido a contenidos usados en la planificación que concuerdan o no con las sugerencias del Real Decreto de mínimos para tercero de ESO.
- Causa del error del contenido no adecuado.
- Otro contenido relacionado con CF-S en EF.
- Relación con las tecnologías de la información y las comunicaciones (TICs), alumnos con necesidades específicas de apoyo educativo (ANEAE), entorno del centro y competencias.

Una última variable fue construida a partir de las anteriores, con el fin de dar respuesta al objetivo (b) del estudio. Consistió en un indicador que sitúa a la planificación analizada en un lugar de entre 0 y 100, siendo el 0 el peor resultado y 100 el mejor. Se consideraron 16 categorías que oscilaron de 0 a 1, computadas como (suma total*100)/16:

- Número de sesiones de fuerza-resistencia, resistencia aeróbica y flexibilidad. Se consideró 0 cuando no se les dedicaba ninguna sesión; 0,5 cuando fueron menos de 14 sesiones en cada UD; y 1 si al menos en una UD contempló 14 o más sesiones para cada cualidad (0-1 en cada cualidad).
- Objetivos formulados como contenidos. Porcentaje respecto al total (está dividido por 100 e invertida la escala para que un número mayor signifique que es mejor; la puntuación osciló de 0-1).
- Objetivos correctos según la concreción. Porcentaje respecto al total (está dividido por 100; la puntuación osciló de 0-1).
- Contempla los diferentes criterios de evaluación del Real Decreto de enseñanzas mínimas de la ESO. Se considera 0 si no contemplaba el criterio correspondiente en ninguna UD y 1 si al menos lo contemplaba en una UD (1 punto por cada criterio; dado que hay 6 criterios, la puntuación total osciló de 0-6).
- Contenidos adecuados a tercero. Porcentaje respecto al total (se consideraron tanto los contenidos reflejados en las UD como en las sesiones y se dividió por 100; la puntuación osciló de 0-1).
- Usa TICs, ANEAE, entorno educativo del centro, y competencias. Se consideró 0 si no contemplaba estos elementos y 1 si al menos los contemplaba en una UD (0-1 en cada elemento).

2.3. Procedimiento

Tras recibir información y prácticas sobre la planificación de la EF en la asignatura de Fundamentos de la Enseñanza durante un semestre académico (60 h), se les solicitó a los profesores en formación inicial una planificación anual para tercero de la ESO. Se establecieron unas condiciones previas que marcaban datos importantes para la planificación como las finalidades educativas del centro, características generales de los alumnos y familias (incluyendo a los ANEAE) o el entorno físico que rodeaba al centro educativo. Los participantes realizaron la planificación durante el segundo semestre de los cursos 2012-13 y 2013-14, con el asesoramiento del profesor de la asignatura en sesiones teóricas, y tutorías particulares en caso de que se solicitara por parte de los participantes para resolver cualquier duda que les surgiera en el proceso.

El análisis de las planificaciones tuvo lugar durante los meses de mayo a julio del curso académico 2013-14. En primer lugar se establecieron las categorías a analizar (ver proceso en el apartado de instrumento), se entrenó al codificador en la consistencia de codificación y en la validez de su análisis, y se configuró, producto de dichas sesiones de entrenamiento, un manual de codificación que guiaría al codificador en el análisis definitivo.

2.4. Análisis estadístico

Se creó una nueva variable a partir de las ya codificadas para hallar el índice de efectividad de la planificación de la CF-S. Posteriormente se realizó un análisis descriptivo (frecuencias y porcentajes) de todas las variables codificadas en las planificaciones. Todos los análisis estadísticos se realizaron mediante el paquete estadístico SPSS versión 20.0 para Windows (IBM® SPSS® Statistics).

3. Resultados

A continuación se presentan en las Tablas 1-8 los resultados descriptivos de las variables analizadas, así como los resultados del índice de calidad de las planificaciones.

Tabla 1. Frecuencia de unidades didácticas totales y de condición física-salud. Trimestre en el que se planifican

Variable Valores	*f*	%
Número de UD totales (N = 180)		
0-5	47	26,1
6-10	98	54,5
11-15	32	17,8
16-20	3	1,8
Número de UD de CF-S (N = 180)		
1	70	38,9
2	45	25,0
3	49	27,2
4	16	8,9
*En qué trimestre se planifican las UD de CF-S (N = 369)**		
Primero	294	79,7
Segundo	46	12,5
Tercero	29	7,9

Nota: CF-S = Bloque de Condición Física y Salud; *f* = Frecuencia; UD = Unidades Didácticas; (*) N se refiere al número de UD de CF-S planificadas por los 180 profesores en formación inicial.

El número de UD totales osciló entre tres y 18. Lo más frecuente fue planificar seis UD en el año (31 profesores en formación inicial). El número de UD de CF-S osciló entre una y cuatro, siendo la moda planificar solo una (70 profesores en formación inicial). La mayor parte de las UD de CF-S se planifican en el primer trimestre, que conlleva necesariamente aplicar refuerzos durante el curso escolar.

Tabla 2. Frecuencia del porcentaje de unidades didácticas de condición física-salud respecto al total. Sesiones de fuerza-resistencia, resistencia aeróbica y flexibilidad respecto al umbral de desarrollo (≥ 14 sesiones). Aplicación de refuerzos intermitentes

Variable Valores	*f*	%
Porcentaje de UD de CF-S respecto al total (N = 180)		
0 hasta 25	95	52,8
26 hasta 50	79	43,9
51 hasta 75	6	3,3
>75	0	0
Porcentaje de sesiones de CF-S respecto al total (N = 179)		
0 hasta 25	62	34,6
26 hasta 50	108	60,4
51 hasta 75	8	4,4
>75	1	0,6
Aplican refuerzo intermitente (N = 180)		

No	164	91,1
Sí	16	8,9
Número de sesiones de resistencia aeróbica (N = 180)		
0-13	168	93,3
14 o más	12	6,7
Número de sesiones de flexibilidad (N = 180)		
0-13	168	93,3
14 o más	12	6,7
Número sesiones de fuerza-resistencia (N = 180)		
0-13	172	95,5
14 o más	8	4,4

Nota: CF-S = Bloque de Condición Física y Salud; *f* = Frecuencia; UD = Unidades Didácticas.

El mayor porcentaje de UD de CF-S se situó entre el 0-50% de las UD totales planificadas, siendo la moda el 25% (26 casos). Un 47,2% de los profesores en formación inicial otorgó más importancia a este bloque que al resto, puesto que superaron el 25% de UD planificadas respecto al total (habiendo cuatro bloques de contenidos en el currículo de EF). Respecto a las sesiones destinadas a CF-S, lo más frecuente fue dedicar el 29,4% de sesiones del curso escolar (22 casos). El 60,4% de los profesores en formación inicial analizados dedicaron entre el 25 y el 50% de las sesiones a la CF-S (contando, no solo las sesiones de fuerza-resistencia, resistencia aeróbica o flexibilidad, sino también por ejemplo la higiene postural o alimentación). Para analizar este resultado de las sesiones dedicadas a la CF-S como bloque de contenido se tuvieron en cuenta todas las temáticas de este bloque.

Solo 12 profesores en formación inicial de los analizados tuvieron un número de sesiones por encima del umbral establecido para poder desarrollar la flexibilidad y la resistencia aeróbica, y solo ocho profesores en el caso de la fuerza-resistencia.

Más del 90% de las planificaciones analizadas no aplicaron refuerzos intermitentes para mantener los incrementos en CF-S. Sin embargo, dentro de los 16 casos que sí lo utilizaron, alguno solo lo aplicó a una cualidad y no a las otras dos, además solo 13 aplicaron cuatro sesiones o más (criterio establecido como eficaz dadas las evidencias científicas) tras un periodo de desentrenamiento de aproximadamente cuatro semanas.

Respecto a los objetivos de las UD (conceptos, procedimientos y actitudes), las actitudes fueron las más olvidadas con el 42,5% de los casos (74 profesores en formación inicial no planificaron ningún objetivo centrado en actitudes), seguidas de los conceptos (14,9%, 26 profesores en formación inicial no planificó ningún objetivo conceptual) y los procedimientos (9,2%, 16 profesores en formación inicial no tuvieron en cuenta procedimientos en la planificación de la CF-S). Lo más frecuente fue redactar en las UD de CF-S un solo objetivo (en el caso de las actitudes) o uno y dos objetivos en el caso de los procedimientos y conceptos.

Se detectaron 63 casos de objetivos mal formulados (sin verbo en infinitivo), mientras que se formularon correctamente 117 casos. Sin embargo desde el punto de vista de la concreción, todos los casos analizados (179 = 100%) tuvieron problemas en este sentido. El 69,3% de los casos analizados tuvieron correctos el 50% de sus objetivos o menos, según la concreción. Los profesores en formación inicial usaron la metodología, las TICs, el entorno del centro educativo y los ANEAE como estrategias de concreción al formular objetivos.

Tabla 3. Frecuencia del cumplimiento de contemplar en la planificación los criterios de evaluación del Real Decreto 1631/2006, de 29 de diciembre, de enseñanzas mínimas para secundaria (N = 180)

Variable	*f*	%
Valores		

CE 1 (relación aparatos-sistemas)		
No	68	37,8
Sí	112	62,2
CE 2.1 (desarrollar fuerza-resistencia)		
No	9	5,0
Sí	171	95,0
CE 2.2 (desarrollar resistencia aeróbica)		
No	8	4,4
Sí	172	95,6
CE 2.3 (desarrollar flexibilidad)		
No	26	14,4
Sí	154	85,6
CE 4 (balance calórico)		
No	57	31,7
Sí	123	68,3
CE 3 (salud postural)		
No	98	54,4
Sí	82	45,6

Nota: CE = Criterio de evaluación del Real Decreto de enseñanzas mínimas de la Ley Orgánica de Educación en tercer curso de Secundaria; *f* = Frecuencia.

El segundo criterio de evaluación que establece desarrollar las cualidades relacionadas con la salud (fuerza-resistencia, resistencia aeróbica y flexibilidad) fue incluido en un alto porcentaje en las planificaciones (85-95%), pero en las tres cualidades se dieron casos que no lo incluyeron. Los criterios de evaluación más olvidados fueron los relacionados con la salud-higiene postural para prevenir lesiones (54,4%), el referido a la relación entre aparatos y sistemas con los efectos que produce la actividad física (37,8%), y el relacionado con la alimentación-balance calórico (31,7%).

Tabla 4. Frecuencia de errores en la formulación de contenidos. Frecuencia por rangos de porcentaje de contenidos adecuados y no adecuados a tercer curso de la ESO

Variable Valores	*f*	%
Contenido formulado como objetivo (N = 180)		
No	144	80,0
Sí	36	20,0
Porcentaje de contenidos adecuados a tercero ESO en unidades didácticas (N = 175)		
0	46	26,3
0,01-30,0	26	14,7
30,01-60,0	54	31,1
60,01-90,0	35	28,8
>90,0	14	8,0
Porcentaje de contenidos no adecuados a tercero ESO en unidades didácticas (N = 175)		
0	13	7,4
0,01-30,0	25	14,3
30,01-60,0	47	27,4
60,01-90,0	43	24,6
>90,0	46	26,3

Nota: ESO = Educación Secundaria Obligatoria; *f* = Frecuencia, número de casos en que se da el valor de la variable.

En 36 casos (20%) se formularon contenidos con verbos en infinitivo, confundiéndolos con objetivos. En 46 casos (26,3%) no se detectó ningún contenido adecuado al curso programado según el Real Decreto de enseñanzas mínimas. El 40% de los casos analizados tenían el 50% o más de los contenidos formulados con errores, como no especificar lo esencial para tercer curso (por ejemplo, nombrar las cualidades físicas sin especificar sus métodos de entrenamiento) o formular contenidos propios de otros cursos diferentes al programado.

Tabla 7. Frecuencia de relación de la CF-S con los demás contenidos de Educación Física, con el uso del entorno, inclusión de ANEAE, TICs y con competencias (N = 180)

Variable Valores	*f*	%
Relaciona CF-S con otros contenidos de Educación Física		
No	120	66,7
Sí	60	33,3
Usa TICs relacionadas con CF-S		
No	105	58,3
Sí	75	41,7
Contempla los ANEAE en CF-S		
No	176	97,8
Sí	4	2,2
Usa el entorno del centro al planificar la CF-S		
No	133	73,9
Sí	47	26,1
Relaciona alguna competencia básica con CF-S		
No	130	72,2
Sí	50	27,8

Nota: ANEAE = Alumnos con necesidades específicas de apoyo educativo; CF-S = Condición física-salud; *f* = Frecuencia; TICs = Tecnologías de la información y comunicaciones.

No se relacionó la CF-S con ningún contenido de EF en el 66,7% de las planificaciones (120 casos). En los casos positivos se relacionaron con los Juegos y Deportes, Expresión Corporal y Actividades en el Medio Natural. Para realizar esta valoración se tomaron los contenidos expresados en las sesiones de las UD de CF-S.

Elementos tan importantes para la planificación en Secundaria como las TICs, los ANEAE, el uso del entorno del centro educativo o las Competencias tuvieron un olvido importante entre las planificaciones analizadas [105 casos en las TICs (58,3%), 130 casos en las Competencias (72,2%), 133 casos en el uso del entorno del centro (73,9%), y los ANEAE con 176 casos (97,8%)].

Tabla 8. Distribución porcentual de las planificaciones en función del índice de efectividad-calidad de la condición física-salud (N = 178)

Variable Valores	*f*	%
Índice inferior a 50		
0-10	1	0,6
11-15	1	0,6
16-20	1	0,6
21-25	2	1,2
26-30	2	1,2
31-35	4	2,4
36-40	8	4,8
41-45	22	13,2
46-50	23	13,8

Índice superior a 50		
51-55	21	12,6
56-60	29	17,4
61-65	16	9,6
66-70	29	17,4
71-75	12	7,2
76-80	3	1,8
>81	4	2,4

Nota: *f* = Frecuencia.

La distribución de las planificaciones según los valores del índice que evalúa su calidad está centrada en los valores medios (entre el 45 y el 70, estando solo 19 de las planificaciones analizadas por encima del valor 70). Sin embargo, 64 planificaciones están por debajo del valor 50.

4. Discusión

Los objetivos de este estudio fueron: (a) analizar las tendencias de los profesores de EF en formación inicial al planificar la CF-S, identificando errores y proponiendo recomendaciones que nos ayuden en el futuro a intervenir desde la propia universidad; y (b) crear un índice que nos indique la calidad-efectividad de una planificación en la programación de la CF-S.

Respecto al primer objetivo, los profesores en formación inicial otorgaron mayor importancia a la CF-S que al resto de contenidos para secundaria, estando en línea según el análisis realizado por Viciana, Salinas, y Cocca (2007), con las recomendaciones que realiza el currículo español. Estos autores, ya determinaron en su análisis de contenido del currículo de secundaria que la CF-S era el contenido más mencionado en el Real Decreto de enseñanzas mínimas para secundaria, y por tanto, dedujeron que igualmente debía ser el planificado y tratado con mayor dedicación por el profesorado (Ministerio de Educación, Cultura y Deporte, 2007).

Sin embargo, esta importancia de la CF-S puede no tener efecto alguno sobre la salud escolar si no se planifica correctamente. Apoyados en los estudios de Mayorga-Vega et al. (2013) y Viciana et al. (2013), se siguió el criterio de aplicar 14 sesiones de intervención como las mínimas necesarias para obtener un aumento en el nivel de CF-S con respecto al nivel inicial (exigido por la legislación española). No se ha encontrado en la literatura científica ningún otro estudio con menor tiempo de intervención en el que los estudiantes hayan incrementado su CF-S respecto al nivel inicial (ya sea fuerza-resistencia, resistencia aeróbica o flexibilidad). Las planificaciones analizadas mostraron deficiencias en este sentido, siendo un escaso número de programaciones las que cumplieron este criterio de eficacia para conseguir el aumento requerido. Por tanto, estos resultados sugieren que planificar la CF-S con intervenciones cortas fracasarán e incumplirán el primer nivel curricular español.

Sorprende el bajo porcentaje de profesores en formación inicial que ha utilizado refuerzos intermitentes en sus planificaciones para mantener los incrementos de CF-S logrados previamente. Mayorga-Vega et al. (2013) aplicaron con éxito cuatro sesiones de refuerzo de la fuerza-resistencia y la resistencia aeróbica en el contexto de la EF escolar para mantener el incremento conseguido inicialmente. En el citado estudio se demostró no solo que era posible el aumento de la CF-S en programas a corto plazo, superando las dificultades del contexto escolar, sino que la aplicación de refuerzos intermitentes para mantenerla también fueron efectivos. Por tanto, aplicar refuerzos intermitentes de cuatro sesiones al menos para mantener la fuerza-resistencia y resistencia aeróbica del alumnado fue tomado como criterio para evaluar la eficacia de las programaciones en este estudio. Intervenciones en la formación inicial para concienciar de la importancia de estos datos son necesarias a la luz de los resultados obtenidos en esta investigación.

Los conceptos, procedimientos y actitudes deben ser tenidos en cuenta en la planificación de cualquier intervención en el contexto escolar (Viciana, 2002). Si analizamos los verbos usados en los criterios de evaluación de CF-S del currículo para tercer curso de la ESO, vemos cómo el Ministerio de Educación está solicitando a los profesores de EF españoles especial atención a los conceptos y procedimientos. Ejemplos de un criterio de evaluación conceptual y otro procedimental serían: "Analizar los efectos beneficiosos y de prevención que el trabajo regular de resistencia aeróbica, de flexibilidad y de fuerza-resistencia suponen para el estado de salud" y "Diseñar y llevar a cabo un plan de trabajo de una cualidad física relacionada con la salud, incrementando el propio nivel inicial, a partir del conocimiento de sistemas y métodos de entrenamiento" (Ministerio de Educación Cultura y Deporte, 2007, pp. 714-715). Los objetivos dedicados a las actitudes, a pesar de ser tomados en cuenta algo menos que los objetivos dirigidos a conceptos y procedimientos, son de gran importancia para la práctica de actividad física y la CF-S de los escolares. Además, su fin último está dirigido a aplicar los aprendizajes de este contenido a la vida cotidiana, así como en el futuro durante la adultez. Dicha práctica de actividad física da lugar a una adherencia necesaria para hacer frente al problema del sedentarismo y prevenir la obesidad, el sobrepeso y enfermedades asociadas, promocionando así una actividad física saludable en los jóvenes (Cocca et al., 2014) que les permita tener una vida adulta igualmente saludable (Ortega et al., 2008).

La exclusión de ciertos elementos del caso práctico a planificar por los profesores en formación inicial, como fueron el recomendado uso de las TICs, la atención a los ANEAE o el desarrollo de las competencias es otro gran problema detectado. En la mayoría de los casos no se tuvieron en cuenta. En la última década especialmente se viene destacando la enorme importancia de la integración de las TICs (Marqués, 2012), de la atención a ANEAE (Arnáiz, 2009; Susinos, 2009) y al desarrollo de las competencias (Molina y Antolín, 2008), siendo cruciales para el desarrollo integral del alumnado en cualquier materia curricular. El problema se agrava si además consideramos que el supuesto práctico a planificar incluía estos elementos como finalidades educativas del centro en el que los profesores en formación inicial desarrollarían la EF.

Respecto al segundo objetivo del estudio relacionado con el índice de calidad/efectividad de la planificación, hay un número considerable de casos por debajo de 50 (36%), lo cual indicaría que dichas planificaciones son deficientes. Además, muchas planificaciones con índices superiores a 50 podrían ser inefectivas para desarrollar la CF-S, ya que no cumplen con los requisitos básicos de duración del programa o aplicación de refuerzos. Esta deficiencia implicaría la ausencia de resultados positivos en la CF-S de los alumnos. ¿Es por tanto la ausencia de salud de los escolares un problema que se inicia en la propia planificación de la EF? Aunque existen otros muchos factores que determinan bajos índices de salud entre los escolares [por ejemplo una inadecuada alimentación o la falta de apoyo de los otros significativos (Cocca et al., 2014)], queda patente la importancia de tomar conciencia de realizar una correcta planificación. De esta concienciación dependerá en buena medida que la escuela contribuya a disminuir este problema social. La formación inicial en las universidades debe ser uno de los motores para esta labor, así como para la formación correcta en competencias de planificación del profesorado de EF hacia la CF-S.

Como perspectiva futura, creemos necesaria la realización de trabajos similares con profesores en servicio, aportando la certeza de que lo hallado en este estudio se aplica posteriormente a las clases de EF. Sin embargo, somos conscientes de la dificultad de acceder a las programaciones de profesores en servicio.

Conclusiones

Los resultados sugieren que los profesores en formación inicial coinciden con el Real Decreto de secundaria al dotar de gran importancia a la CF-S, pero no son congruentes en su planificación. Aplican un número insuficiente de sesiones de intervención que imposibilita el incremento perseguido de CF-S en los escolares. Asimismo, el número de profesores en formación inicial que utiliza refuerzos intermitentes es realmente bajo, demostrándose que, si en el mejor de los casos la CF-S se incrementara, ésta se perdería con el paso de varias semanas, llegando los alumnos a final del año con los valores que comenzaron. Se precisa por tanto de una exhaustiva labor de concienciación en la formación inicial para inculcar esta correcta planificación y metodología en el ámbito de la CF-S. Implementar y evaluar en la fase de prácticas de la formación inicial planes correctos de incremento de la CF-S es necesario. Se precisa intervenir específicamente en las actitudes (cruciales en la actividad física y prevención del sedentarismo) incluyéndolas en el aprendizaje del alumno y haciendo partícipes a las familias en el control de los comportamientos en la vida diaria.

Adaptar la planificación de EF a las finalidades educativas del centro es fundamental en la formación inicial. Ejemplos válidos serían el análisis de casos prácticos, estudios colaborativos con familias y comunidades, así como usar el entorno del centro para el desarrollo de la CF-S.

Consideramos de gran importancia estos resultados que confirman, por primera vez, que la CF-S no es planificada correctamente por los profesores en formación inicial, pudiéndose acentuar el consabido problema de salud escolar si estos profesores accedieran al campo laboral.

5. Referencias bibliográficas

Anguera, M. T. (1988). *Observación en la Escuela*. Barcelona: Graó.

Arnáiz, P. (2009). Análisis de las medidas de atención a la diversidad en la Educación Secundaria Obligatoria. *Revista de Educación, 349*, 203-223.

Blair, S. N. (2009). Physical inactivity: The biggest public health problem of the 21st century. *British Journal of Sports Medicine*, *43*(1), 1-2.

Cocca, A.; Liukkonen, J.; Mayorga-Vega, D., & Viciana, J. (2014). Health-related physical activity levels in Spanish youth and young adults. *Perceptual and motor skills, 118*(1), 247-260. doi: 10.2466/10.06.PMS.118k16w1

Emmer, E. T. (1986). Academic activities and task in first-year teachers' classes. *Teaching and Teacher Education*, *2*, 229-244.

Hallal, P. C.; Victora, C. G.; Azevedo, M. R., & Wells, J. C. K. (2006). Adolescent physical activity and health: A systematic review. *Sports Medicine*, *36*, 1019-1030.

Le Ny, J. F. (1980). *Le conditionnement et l'apprentissage*. París: PUF.

Marqués, P. (2012). Impacto de las TIC en la educación: funciones y limitaciones. *3ciencias*, *4*, 1-15.

Mayorga-Vega, D.; Viciana, J., & Cocca, A. (2013). Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics, 37*, 153-160. doi: 10.2478/hukin-2013-0036

Merino-Marban, R.; Mayorga-Vega, D.; Fernandez-Rodríguez, E.; Vera-Estrada, F., & Viciana, J. (2015). Effect of a physical education-based stretching program on hamstring extensibility and its posterior reduction in elementary schoolchildren. *European Physical Education Review, 21*(1), 83-92.

Ministerio de Educación, Cultura y Deporte (2007). *Real Decreto 1631/2006 de Enseñanzas Mínimas para Educación Secundaria Obligatoria*. Madrid: BOE número 5 de 7 de enero de 2007.

Ministerio de Educación, Cultura y Deporte (2014). *Real Decreto 126/2014 por el que se establece el currículo básico de la Educación Primaria*. Madrid: BOE número 52 de 1 de marzo de 2014.

Molina, J. P., & Antolín, L. (2008). Las competencias básicas en educación física: Una valoración crítica. *Cultura, ciencia y deporte, 3*(8), 81-86.

National Association for Sport and Physical Education (2004). *Moving into the future: National Standards for physical education* (2nd ed.). Reston, VA: Author.

Organización Mundial de la Salud (2011). *Global Recommendation on Physical Activity for Health*. Tomado el 22 de octubre de 2014 de http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf.

Ortega, F. B.; Ruiz, J. R.; Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, *32*(1), 1-11. doi: 10.1038/sj.ijo.0803774

Pérez-Pastur, R. (2010). Unidad didáctica sobre las capacidades físicas básicas. *Eduinnova*, *25*, 95-106.

Placek, J. H. (1984). A multi-case study of teacher planning in physical education. *Journal of teaching in physical education*, *4*(1), 39-49.

Quintana, A. (2006). Metodología de investigación científica cualitativa. En A. Quintana y W. Montgomery (Eds.), *Psicología: tópicos de actualidad* (47-84). Lima, Perú: UNMSM.

Robles, J.; Giménez, F. J., & Abad, M. T. (2010). Motives for physical education teachers to select sport contents in secondary school. *Retos, 18*(2), 5−8.

Rodríguez, P. L.; Santonja F. M.; López-Miñarro, P. A.; Sáinz de Baranda, P., & Yuste, J. L. (2008). Effect of physical education stretching program on sit-and-reach score in schoolchildren. *Science & Sports*, *23*(3), 170-175.

Ruiz, J. R.; Castro-Piñero, J.; Artero, E. G.; Ortega, F. B.; Sjöström, M.; Suni, J., & Castillo, M. J. (2009). Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine*, *43*(12), 909-923.

Salinas, F. (2011). *Conceptions of pre-service physical education teachers about planning* (Unpublished Thesis Dissertation). University of Granada, Spain.

Susinos, T. (2009). Escuchar para compartir. Reconociendo la autoridad del alumnado en el proyecto de una escuela inclusiva. *Revista de Educación, 349*, 119-136.

Tremblay, M. (2014). 2014 global summit on the physical activity of children. *Journal of Physical Activity and Health*, 11(Supp 1), S1-S2. doi: 10.1123/jpah.2014-0182

Viciana, J. (2002). *Planificar en Educación Física*. Barcelona, Spain: Inde.

Viciana, J.; Lozano, L.; Cocca, A., & Mayorga-Vega, D. (2012). Influence of the organizational system on motor engagement time in physical education on high school students. *Procedia-Social and Behavioral Sciences, 69*, 1160-1167.

Viciana, J., Mayorga-Vega, D., & Cocca, A. (2013). Effects of a maintenance resistance training program on muscular strength in schoolchildren. *Kinesiology*, *45*(1), 82-91.

Viciana, J.; Mayorga-Vega, D., & Cocca, A. (2014). Modelo de aprendizaje exitoso en educación física y su mantenimiento. Estudio del efecto del refuerzo intermitente sobre la condición física. *Revista iberoamericana de psicología del ejercicio y el deporte*, *9*(1), 155-171.

Viciana, J.; Salinas, F., & Cocca, A. (2007). Análisis de contenido comparativo del primer nivel curricular de Educación Física en Secundaria. *Revista de curriculum y formación del profesorado*, *11*(2), 257-271.

Viciana, J.; Salinas, F., & Cocca, A. (2008). Effects of an aerobic-capacity based program on cardiovascular risk factors in Spanish schools. *Circulation*, *117*(11), e221.

Viciana, J., & Zabala, M. (2004). Un estudio descriptivo sobre cómo planifican los profesores de Educación Física. En M. González, A. Sánchez y J. Gómez (Eds.) *International Congress of AIESEP 2002. Preparación profesional y necesidades sociales* (pp. 732-739). A Coruña, Spain: AIESEP.

2. PHYSICAL EDUCATION-BASED PLANNING FOR DEVELOPING AND MAINTAINING STUDENTS' HEALTH-RELATED PHYSICAL FITNESS LEVELS (PAPERS III-IV)

Paper III

INNOVATIVE TEACHING UNITS APPLIED TO PHYSICAL EDUCATION. CHANGING THE CURRICULUM MANAGEMENT FOR AUTHENTIC OUTCOMES

Viciana, J., & Mayorga-Vega, D.

Kinesiology

Accepted for publication

Kinesiology

International Journal of Fundamental and Applied Kinesiology

Horvaćanski zavoj 15, 10 000 Zagreb, Croatia

Phone: 385 1 3658 742, Fax: 385 1 36 34 146, E-mail: kinesiology.office@kif.hr

Zagreb, October 22nd, 2015

Daniel Mayorga-Vega
ID number: -74879318N-
Department of Physical Education and Sport
University of Granada, Granada, Spain

Hereby we confirm that the paper No 1217 under the title "**INNOVATIVE TEACHING UNITS APPLIED TO PHYSICAL EDUCATION. CHANGING THE CURRICULUM MANAGEMENT FOR AUTHENTIC OUTCOMES**" written by Jesús Viciana and Daniel Mayorga-Vega is accepted for publication in the journal **Kinesiology**. It shall be published in December 2015, in the volume 47, issue 2 of the journal **Kinesiology**.

Sincerely,

Editor-in-Chief:

Prof. Dragan Milanović, PhD

INNOVATIVE TEACHING UNITS APPLIED TO PHYSICAL EDUCATION CHANGING THE CURRICULUM MANAGEMENT FOR AUTHENTIC OUTCOMES

Jesús Viciana and Daniel Mayorga-Vega

Department of Physical Education and Sport, University of Granada, Granada, Spain

Correspondence to:

Prof. Jesús Viciana, Ph.D.

Department of Physical Education and Sport, University of Granada

Ctra. Alfacar s/n, 18011, Spain

Phone: +34 958246641

Fax: +34 958244369

E-mail: jviciana@ugr.es

Submitted: September 20, 2014

Accepted: April 17, 2015

Abstract

The aim of this article is to reflect on the original attributes that define planning in order to reformulate the concept of teaching units in physical education (PE), providing teachers with new structures of planning that allow them to achieve authentic outcomes. Some issues related to students' learning are raised as regards to the duration of a traditional teaching unit, as well as the total time amount and allocation dedicated to student's learning. Based on previous psychological and learning theories, as well as on in-service teachers' experience, four proposals of innovative teaching units are provided. Providing opportunities for significant and meaningful learning level in physical education is not an easy task for teachers. The main objective of the proposals is to provide students with substantial learning experiences and to accomplish authentic outcomes by planning PE classes. Intermittent, alternated, irregular, and reinforced teaching units are analysed. The proposals focus on new approaches to time allocation in PE, and on a renovated concept of unit planning and curriculum design. Several applications for each type of teaching unit are shown in order to highlight PE teachers' opportunities to innovate their planning. The proposals of intermittent, alternated, irregular, and reinforced teaching units could guide teachers through planning of PE based on related learning and may provide an effective tool for PE planning to overcome learning deficiencies attributed to traditional teaching units.

Keywords: *learning processes, physical education teachers, traditional teaching unit, curriculum organization, length of the unit*

Introduction

The evolution of planning in education was initially marked by the United Nations Educational, Scientific, and Cultural Organization (UNESCO, 1962) recommendations, whereby planning was associated with the evolution of society (attending to pupils' future needs more than the present requirements), thus giving rise to short- and long-term planning concepts applied to educational stages: Recommendation number 18 referred to "institutional planning should be directed to ensuring maximum efficiency and flexibility in use" (UNESCO, 1974:8). As a result of this recommendation, planning in education has been theoretically associated with the characteristics of flexibility, efficiency, and based on objectives (Viciana, 2002). All nations, depending on their own educational policy, socio-cultural and socio-economic characteristics, political and ideological factors as well as external influences, have developed their standard guidelines or educational aims from which the districts (communities/neighbourhoods) and school centres program their educational goals (with regard to the general profile of their students, and their community characteristics). Following these educational goals planned by centres, teachers programme physical education (PE) adapting their teaching to the students (Kelly & Melograno, 2004; Viciana, 2002).

A teaching unit in Physical Education

In contrast to the original planning characteristics of flexibility, efficiency and being objective-based (Siedentop, 1998; Viciana, 2002), one of the most used planning structures in the subject of PE has been developed around the concept of a teaching unit (TU) or unit of instruction (Kelly & Melograno, 2004; Piéron, 1992). Teaching unit is understood as a closed and traditional concept associated with several problems. First, TU has been defined as a part of the annual plan made by physical educators (Pesquie, 1988; Piéron, 1992), a period of time that teachers only fill with several subject matters. The concept of "covering the curriculum"

(Siedentop & Tanehill, 2000), where teachers feel the necessity of treating a wide range of contents during a year, has caused ineffective short TUs (Himberg, Hutchinson, & Roussell, 2003; Robles, Giménez, & Abad, 2010). In this line, for example, pre-service and novice PE teachers usually design very short TUs (two or four weeks, twice a week) with the main objective to make students develop their physical fitness (Pérez-Pastur, 2010).

Nevertheless, recent studies express concern that the length of TU is crucial in learning (Araújo, Mesquita, & Hastie, 2014). Secondly, TU has been conceived as a particular number of classes consecutively conducted in the school centre to attain the set objective (or a group of them), and in the best case scenario, teachers consider the necessary learning-time (based on their own experience or taking the reference of previous studies) to achieve it/them (Viciana, Salinas, & Cocca, 2008). However, although many studies have been effective in increasing the levels of students' learning in PE settings (i.e. fitness, motor skills, tactical understanding in games) (Ardoy, et al., 2010; Westendorp, et al., 2014; Gray & Sproule, 2011); unfortunately none of them mentioned for how long the achieved learning level was sustained. However, recent research has started to analyse this issue (Mesquita, Farias, & Hastie, 2012; Pereira, et al., 2015). Lastly, the traditional concept of TU raises the third issue. Traditional TUs have usually been based on the achievement of isolated objectives and isolated contents acquisition (Kelly & Melograno, 2004; Piéron, 1992; Siedentop & Tanehill, 2000) failing to establish the relationships between those objectives and the rest of the subject, or the situational perspective of the learning (Smith, 2011).

From the traditional point of view, the most important aspect of TU in PE has been its duration that is represented by the number of classes (minutes) needed to achieve the objective (Delaunay & Pineau, 1989; Viciana, 2002). This duration includes the overall learning time spent and the time students needed to learn something as a part of it (Van der Mars, 2006). Based on the Carroll's model of school learning (1963), saying that the degree of learning could be represented by a function, many authors modified the model in order to complete the variables that ensured an effective teaching process (Cruickshank, 1985; Huitt, 2006; Van der Mars, 2006). In the numerator of the model's function was the "time spent learning" (that will be the academic learning time in the future: it is composed of the allocated time or opportunity to learn, and of the students time spent in learning tasks or in perseverance). The denominator was "time needed to learn" (composed of students' aptitude for learning the contents, their ability to understand the instruction, and the quality of that instruction). This model, modified by successive authors, has contributed step-by-step to the total comprehension of a good framework for students' motor learning in PE (mainly in behavioural skills), and also added new variables for including students' social skills in the model of the teaching/learning process (McIlrath & Huitt, 1995).

However, based on the original UNESCO's conception of planning (i.e. recognising the attributes of flexibility and efficiency and based on educational objectives), and on an innovative TU conception (beyond the traditional point of view), a new variable could be taken into account for improving those models and students' learning: this is the *innovative distribution of TU time* throughout an annual plan in PE. This new element could be included in the numerator of the formula of the school model for learning (and the subsequent modified models) in order to improve three important elements regarding learning (but maintaining the time spent in learning and the time needed to learn constant): (a) provide situational learning in PE; (b) improve the relational cognitive-behavioural learning between PE contents; and (c) develop and maintain the cognitive and behavioural learning level already achieved. This new perspective of the time-learning distribution along the academic year could facilitate teachers to attain authentic outcomes, based on meaningful learning, instead of being contented with easily forgettable simple goals (Mayorga-Vega, Merino-Marban, Vera-Estrada, & Viciana, 2014).

Consequently, the purpose of this article is to highlight the importance of innovative forms in the distribution of time planning in PE along the annual plan, and to provide new planning possibilities (foreseeing particular divisions and distributions of time for learning) for teachers in order to solve some planning-related learning issues in PE.

Planning-related learning issues in Physical Education

With regard to curriculum and planning errors, Siedentop and Tanehill (2000) commented that one major culprit was the continued widespread use of the short-unit, multi-activity curriculum model that many referred to as the "smorgasbord" curriculum. Because PE has many contents and goals, some teachers plan the curriculum as a series of short activity units, with a few classes of isolated basic skills practice. This approach could lead to continued students' efforts to learn in those short periods of teaching, since students are continuously in a discovery phase, and are not able to achieve the learning with so limited opportunities for successful practice (Viciana, 2002). Consequently, boredom emerges in students and the failure of learning is certain. Although simple objectives could be achieved after a short TU, depending on the nature of learning pursued, authentic outcomes commonly fail. Moreover, learning is usually isolated, with no relationship to the whole subject matter (Zhang, et al., 2014). However, recent studies regarding other curricular models such as the sport education model (usually centered in social interactions) addressed these issues, showing importance, and complexity of the factors related to learning (Araújo, et al., 2014).

Regarding the time of learning opportunities, Van der Mars (2006) represented a multi-level scheme of the influences that determine the student's time/opportunity to learn in PE. This scheme includes: (a) a national and state level legislation that regulates the recommended minutes of PE per week; (b) a district level legislation that sets the length/calendar of the lesson; (c) school level regulations that decide on the length of TU and preferred activities; and (d) program/class level regulations that define the time of lesson, allocated time, engagement time, moderate-to-vigorous physical activity time, and academic learning time in PE.

For instance, in Spain the Ministry of Education has established two one-hour lessons per week (≅65-68 hours/lessons per year) for PE in the curriculum of elementary and high school, which, in reality, are reduced to 45-50 min each due to the transition and displacement between classrooms (Ministry of Education, 2006). This aspect of the current Spanish educational law is widely criticised by PE teachers (Viciana, 2002). This similarly limited time assigned to PE is shared by numerous European countries, where the average time allocation ranges from 53-81 hours/year (European Commission/EACEA/Eurydice, 2013) and, consequently, all of them share this learning-regarded problem and its consequences. It supposes 65 hours per year approximately [varying among the countries, i.e. in primary education the minimum recommendation ranges from 37 hours in Ireland to 108 hours in France (European Commission/EACEA/Eurydice, 2013)] with a great number of contents to deliver and objectives to achieve. Despite several studies demonstrating that PE teachers could increase, for instance, students' physical fitness within the traditional design and time allocation (Mayorga-Vega, Viciana, Cocca, & Rueda, 2012; Viciana, et al., 2008), many PE teachers still consider that it is impossible with only two lessons a week (Salinas, 2011; Corbin, et al., 2014).

Most authors reveal their concern about the time they need to achieve an objective in PE. Some authors have observed that it is necessary to deliver a great number of lessons, because a reduced number tends to situate students at a discovery level, without enough time to produce sustained effects on learning (Seners, 2001; Siedentop & Tanehill, 2000). Hébrard (1986) specified that 10 lessons were insufficient for reaching the stage of learning in PE and that students would probably forget the learning gained. Himberg et al. (2003) suggested that the first change for increasing the effectiveness of the curriculum would be to lengthen teaching units from two or three weeks to four to eight weeks. All studies regarding time/opportunity and learning in the PE research line support the link between time and learning, considering that proper instructions are provided during that time (such as feedback, presentation of the task or motor problem to be solved by students, or the style of teaching) (Van der Mars, 2006; Viciana, Lozano, Cocca, & Mayorga-Vega, 2012), but none of them has treated innovative distribution of TU time that could facilitate comprehensive learning in PE, instead of considering the TU as an accumulative number of lessons consecutively delivered, and usually conceived as a short-term planning.

The proposals of innovative TUs below address those issues, mainly concentrated in those countries (and some states of the USA) where the PE schedule is composed of two (or three) lessons a week, showing different opportunities to organise learning/teaching periods planned by PE teachers in the annual curriculum. However, some limitations such as local regulations, school contexts, or students' characteristics could affect the application of the proposals in this article.

New structures for planning teaching units in Physical Education

The educational objective is the main element of the curriculum that guides the teaching-learning process and planning (Kelly & Melograno, 2004). It is defined as a statement of instructional intent that specifies what knowledge, attitudes, and behaviours are meant to be learned (in fitness, motor, cognitive, or social domains) (Siedentop & Tanehill, 2000). An outcome is a description of what a student will know and be able to do as the result of participating in the activities of the program. Outcomes could be conceived as "ends", and curriculum objectives as "means" (Lambert, 1996). However, the adjective of *authentic* comes from the "authentic pedagogy theory" of Newmann, Marks, and Gamoran (1995) that has been assigned to outcomes in order to specify the contextual performance where skill, knowledge, or social strategy will be used by the students. The characteristics of meaningful learning (connection with previous knowledge), students' engagement and analysis (psychological involvement of students in their learning), and connection to the world (application of the information and performance to other real sport and physical activity contexts in the students' life) are related to this authentic pedagogy theory, and they constitute the theoretical basis of the following units of planning proposals. Newmann (1992) defined student engagement in an educational setting as the student's psychological investment in an effort focused on situated learning. Situated learning, in contrast with many classroom activities that are out of context, refers to the deliberate situated performances for learning, developed in authentic contexts (with social interactions, collaboration and performed in situations that would normally involve that knowledge).

Proposal I: Intermittent teaching unit

First proposal of innovative TU is called the "intermittent teaching unit". This proposal is based on the frequency of practices distributed during a year or a particular period of time, thus changing the concept of a conventional TU and connecting students' learning with other knowledge of the subject of PE.

Insert Figure 1

In Figure 1 the traditional TU is shown in comparison with the intermittent TU (shaded areas). The intermittent TU is composed of many lessons' pieces of time (from 5 to 15 min) that allow PE teachers to develop a particular content matched to the rest of the physical activity developed during the lessons that compose this intermittent TU. This conception of TU allows teachers to carry out applied lessons instead of artificial lessons based on isolated and decontextualized contents. The intermittent TU could be applied around ten minutes in the introductory or in the cool down phase (first or final parts of lessons). Several studies with variable durations, ranging between nine weeks to one academic year, have carried out different examples of this situational perspective of the development of flexibility in elementary and secondary school settings (Constantino, Aires, & Ramos, 2012; Sánchez Rivas, Mayorga-Vega, Fernández Rodríguez, & Merino-Marban, 2014).

The application of the intermittent TU is centred on some learning and contents that PE teachers usually develop day by day (i.e. the warm-up, methods of training in flexibility, muscular relaxation methods, prevention of injuries, etc.). It seems that these subject matters should not be developed during a traditional TU, because if they are concentrated in several consecutive PE sessions it could be boring and ineffective for students' learning, and the performance would be decontextualized. Moreover, the application of this learning to several physical activities is unusually linked to other subject matters, but isolated.

It could seem that the total amount of time allocated to this intermittent TU would be less than the time of a traditional TU, because only very short periods of time of each lesson are allocated. However, when all these short periods are summed, then it gives the same or more learning opportunities for students than a traditional TU. Instead of traditional lessons, this innovative TU organises the phases dedicated to learning's advance in groups of a particular number of lessons (i.e. 10 minutes during several periods of 5-10 lessons). In each phase of lessons, the content is organised progressively in order to facilitate the students' understanding and participation (see examples in Figure 2).

Insert Figure 2

The number and duration of the phases that compose the total process is adaptable, depending on the objectives programmed by PE teachers. In Figure 2, there are three different examples of objectives centred on students' autonomy in warming-up, on the development of several kinds of flexibility training methods and their application to specific physical activities, and on the attitudinal and cognitive learning related to alimentation and caloric balance.

Proposal II: Alternated teaching units

Alternated TUs allow students to link the learning developed by both TUs (the objectives of the two TUs alternated), thus allowing students to understand the fundamentals that connect them (Figure 3). The example of Figure 3 is based on the teaching games for understanding developed by Bunker and Thorpe (1982), although many other applications could be incorporated into these alternated TUs (Figure 4). The main applications of the alternated TUs to PE are: (a) to relate concepts and behavioural skills belonging to the PE syllabus; and (b) to connect the learning developed in classroom to the students' life. The transfer between concepts and learning has been previously verified in secondary school, in motor skills as well as in team sports (Román, Miranda, Martínez, Martínez, & Viciana, 2007; Yáñez & Castejón, 2011).

Insert Figure 3

Although the transfer and relational learning can be applied between two or more tasks in the same PE lesson, often the learning that PE teachers are interested in is developed within different contexts (e.g. indoor-outdoor), or each one needs different materials and organisations that make it difficult to put into practice both types of tasks in the same lesson (e.g. skating-skiing). In these cases, the alternated TUs are a good solution for promoting significant learning in PE. Moreover, some school centres also have a shared schedule by two PE teachers in the same hour of the same day of the week that produce the problem of sharing the gym. Alternating TUs teachers could solve this problem, teaching one day an indoor content and the other day of the week an outdoor content, thus facilitating the sharing of the available gym between PE teachers. Nevertheless, some limitations such as adverse weather conditions could affect the feasibility of this proposal.

Insert Figure 4

Teaching alternated lessons, as in the examples of Figure 4, allows teachers to relate the learning, make them aware of these relationships between contents and methods in PE, and help them in reaching authentic and meaningful outcomes.

Proposal III: Irregular teaching unit

From using the characteristics of planning flexibility and dynamism at its maximum level, without forgetting the main propose of a TU, that is to reach the objective and authentic outcomes, arises the irregular TU that provides a wide range of possibilities in PE. The irregular TU means to distribute the lessons that compose it over the whole academic year with regard to the main centre of interest pursued by the PE teacher. Thus, many possibilities emerge from this perspective applied to several curriculum models (see example within sport education model in Figure 5).

Insert Figure 5

The amount or time related to learning opportunities is variable and depends on the number of lessons used and the optional use of the extracurricular time (depending on the maturity and autonomy of the students). In Figure 5 the irregular TU uses recess time in order to develop the competition phase, “the season”, in a sport education model. Moreover, the use of extra-curricular time for learning in an irregular TU allows PE teachers to collaborate with students’ families in achieving authentic outcomes and performances applied to life (e.g. irregular TU centred on the daily calculation of caloric balance needs collaboration with the families, or combining PE lessons with physical activity carried out during the weekend with families’ collaboration in order to increase the students’ fitness level). The increment of active time for learning in extra-curricular periods (i.e. after school time, recesses, weekends, holidays) allows teachers to pursue many important objectives (e.g. improve the alimentation, improve the daily physical activity levels, reduce sedentary habits, increase the students’ autonomy). For instance, if students have the required maturity and responsibility, teachers could use extra-curricular time aimed at increasing physical activity during leisure time, thus developing their physical fitness. This could be performed by means of providing the students with the methods of physical fitness development that are characteristic for PE classes, and delegating the responsibility of their development using an individualized style of teaching (individualized programs performed in couples).

Proposal IV: Reinforced teaching unit

The reinforced TU deals with the quantity of practice needed to obtain an increase in a particular learning experience and its reinforcement along the academic course in PE in order to avoid the loss of the learning attained. The length of the unit needs to be foreseen according to the objective pursued and revised according to previous experiences carried out in similar contexts. Delivering extra practice after learning (overlearning) is a crucial factor to provide successful practice for students.

A large problem related to physical fitness, motor skills, and tactical learning is the expected decrease after a period of detraining. Several authors confirm that, for instance, after five to 12 weeks of detraining children lose a significant part of their physical fitness gains obtained (Da Fontoura, Schneider, & Meyer, 2004; Ingle, Sleap, & Tolfrey, 2006; Mayorga-Vega, et al., 2014). Therefore, applying intermittent reinforcements in order to maintain the gain obtained during the initial period is an inherent contribution of this reinforced TU (Figure 6). The intervals of detraining could be used by PE teachers to develop other contents of PE or to consider the holiday periods as detraining. It is necessary to clarify that the instructional approach and the content developed in the first intervention program could affect the detraining period (and the decrease or increase of the learning achieved). Recent studies demonstrated an increase in the levels of students' learning after a detraining period in PE, due to the voluntary involvement in practising physical activity during their leisure time when they have been previously motivated by the teacher and content (i.e. sports) (Mesquita, et al., 2012; Pereira, et al., 2015).

Insert Figure 6

It is important to know that the intermittent reinforcement will have a particular design of activities depending on the learning pursued. For instance, if students' cardiorespiratory endurance during the initial period is to be developed, the design of the tasks during the intermittent reinforcement will include a moderate-to-vigorous intensity performance of generic movements, with a high number of muscle groups involved, as well as tasks with continuous movements and with enough practice time. It is not important whether these tasks are developed through sports or dance, etc., if they follow an adequate methodology. Teachers could apply the intermittent reinforcement during several lessons consecutively delivered or alternating lessons between other contents. The moment for applying the reinforcements will depend on the learning contents (endurance, muscular strength, tactical sport, etc.), and always before students revert to the baseline levels. The study carried out by Mayorga-Vega, Viciana, and Cocca (2013) has confirmed the effectiveness of the intermittent reinforcement in PE setting on physical fitness, creating the fundamentals of the reinforced TUs.

Discussion and conclusions

From the very begining of educational planning, the characteristics of flexibility, efficiency, and being based on objectives were central to this concept. The current contribution has been based on the theory of authentic pedagogy (Newmann, et al., 1995), and proposes new approaches to planning teaching units that comply with the original attributes of planning reccomended by UNESCO (1962). First, flexibility is one of the main characteristics of the developed proposals that break with the traditional concept of TU composed of sessions consecutively delivered, and provide a great variety of allocation and distribution of the time dedicated to learning. Efficiency in students' learning is the aim of the flexibility applied to these proposals of curriculum design, making possible a situational and meaningful learning, as well as retention of the learning achieved during the academic course. And finally, objective-based planning is the motor of the renovated concept of the teaching unit because these proposals have been designed in the form of an educational objective (searching for authentic experiences in PE).

The use of extended or longer units has been a common factor utilised in all proposals throughout this article (with variations according to each of the innovative unit proposed), as suggested by literature (Himberg, et al., 2003; Viciana, et al., 2012): (a) intermittent units have used the concept of pieces of lessons over the whole year in order to provide a higher time-opportunity to learn for students; (b) alternated units have added their lessons (teacher efforts) in order to provide more time to learn related concepts and behaviours; (c) irregular units have used the extra-curricular time to increase the total amount of learning time, and at the same time, the unconnected distribution of its lessons along the months and semesters provide an intermittent stimulus focused on the same objective in order to provide situated learning and authentic outcomes in PE; and (d) reinforced units have used three aspects related to learning-time in PE. The first aspect was to extend the unit as far as other previous studies did in order to assure a certain guarantee of success. The second aspect was the applied overlearning that extends the successful practice for learning, and higher motivation caused by successful practice for students. The third aspect related to learning-time was the application of the intermittent reinforcements in order to maintain the learning previously acquired during the initial period, and to avoid the loss of the acquisition.

As Harnischfeger and Wiley (1985) stated, the active learning time variable is mainly influenced by policy decisions at multiple levels (country, state, district, community, and educational centre). Examples of this influence have been shown by the “block scheduling” innovation made in the USA. This block scheduling consists of longer segments of time for learning provided by increasing the length of a traditional secondary class period (Hackman, Hecht, Harmston, Pliska, & Ziomek, 2001). Models such us 4 x 4 (“four by four”, four classes per semester), or A-B ("alternating day", classes met every other day for 90 min for an entire school year) have been applied, thus increasing the time for learning and obtaining positive results (Bryant & Claxton, 1996; Shortt & Thayer, 1999; Stader, 2001). However, because scheduling modifications depend mostly on external decisions, the treatment of the students' learning time has been utilised in the proposals developed in this article from the perspective of the PE teacher, providing several strategies to increase it merely with the tool of planning. The proposed intermittent, alternated, irregular, and reinforced TUs could guide teachers for planning relational PE contents; facilitate the consecution of situated and significant cognitive and behavioural learning, maintaining it over the academic year; and provide teachers with a planning tool for avoiding the usual loss of learning.

We cannot forget that in the development of this manuscript the meditational processes that occur within the classroom's relationships, between the teacher and student, and among students (McIlrath & Huitt, 1995). Variables such as teacher expectations, feedback (its quantity and quality), interactions provided by the tasks, communication between teachers and students, communication among students, autonomy support, or class size, influence the learning outcomes. Therefore, just as a boring or an unstructured curriculum is destined for failure, an exciting and relevant curriculum could also fail if it is taught ineffectively (Siedentop & Tanehill, 2000). As Silverman, Woods, and Subramaniam (1999) stated, many variables interact for learning to occur. In this article, the authors have pointed to several points as guidance for PE teachers, helping them plan with the innovative structures of the TU, but the rest of the variables that interact in the students' learning process need to be taken into account and carefully designed. These variables may cause a limitation for the application of any of the proposals presented in this manuscript. Teachers need to consider all these variables and decide in which cases the proposals of this article could be limited, depending on the context in which they are to be applied [e.g. a lack of students' maturity or autonomy could cause a failure in the application of irregular teaching units that use extracurricular time (i.e. recess, weekend), due to the absence of a teacher during that time]. Future research lines could be developed in experiencing the proposals provided in this article, verifying their effectiveness, the difficulties according to the nature of the learning and the PE content applied (i.e. sports, physical fitness, dance), as well as in different school contexts.

Using the original characteristics of planning flexibility, dynamically and based on objectives, as mentioned in the introduction section, the proposals of this article have been supported in a new concept of teaching unit, modifying it from the traditional structure of a particular number of sessions consecutively delivered to a new distribution of time according to the educational objective, and achieving the maximum efficiency of learning based on time allocation.

References

Araújo, R., Mesquita, I., & Hastie, P. (2014). Review of the status of learning in research on Sport Education: Future research and practice. *Journal of Sports Science and Medicine, 13*(4), 846-858.

Ardoy, D.N., Fernández-Rodríguez, J.M., Chillón, P., Artero, E.G., España-Romero, V., Jiménez, D.,... Ortega, F.B. (2010). Physical fitness enhancement through education, EDUFIT study: Background, design, methodology and dropout analysis. *Revista Española de Salud Pública*, *84*(2), 151-168.

Bryant, J., & Claxton, D. (1996). Physical education and the four-by-four schedule. *The Physical Educator*, *53*(4), 203-209.

Bunker, D., & Thorpe, R. (1982). A model for teaching of games in secondary schools. *Bulletin of Physical Education, 18*(1), 5-8.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*(8), 723-733.

Constantino, D.H., Aires, G., & Ramos, A. (2012). Chronic effect of static stretching performed during warm-up on flexibility in children. *Brazilian Journal of Kinantropometry, 14*(3), 296-304.

Corbin, C.B., Welk, G.J., Richardson, C., Vowel, C., Lambdin, D., & Wikgren, S. (2014). Youth physical fitness: Ten key concepts. *The Journal of Physical Education, Recreation & Dance, 85*(2), 24-31.

Cruickshank, D. (1985). Profile of an effective teacher. *Educational Horizons, 64*, 90−92.

Da Fontoura, A.S., Schneider, P., & Meyer, F. (2004). Effect of the muscular strength detraining in prepubertal boys. *Revista Brasileira Medicina du Esporte, 10*(4), 285-288.

Delaunay, M., & Pineau, C. (1989). Un programme-la lechón-le cycle en EPS. [A lesson program of the cycle of physical education. In French.]. *Revue EPS, 217*, 49-52.

European Commission/EACEA/Eurydice. (2013). *Physical education and sport at school in Europe-Eurydice Report*. Luxembourg: Publications Office of the European Union.

Gray, S., & Sproule, J. (2011). Developing pupils' performance in team invasion games. *Physical Education and Sport Pedagogy, 16*(1), 15-32.

Hackman, D.G., Hecht, J.E., Harmston, M.T., Pliska, A.M., & Ziomek, R.I. (2001). *Secondary school scheduling models: How do types of models compare to ACT scores?* Poster session presented at the annual meeting of the American Educational Research Association, Seattle. (ERIC Document Reproduction Service Nº ED452230).

Harnischfeger, A., & Wiley, D.E. (1985). Origins of active learning time. In C.W. Fisher & D.C. Berliner (Eds.), *Perspectives on instructional time* (pp. 133-156). New York, NY: Longman.

Hébrard, A. (1986). *L'Èducation physique et sportive. Reflexions et perspectives.* [Physical education and sport. Reflections and perspectives. In French.]. Paris, France: Coedition revue STAPS et Éditions Revue EPS.

Himberg, C., Hutchinson, G.E., & Roussell, J. (2003). *Teaching secondary physical education. Preparing adolescents to be active for life*. Champaign, IL: Human Kinetics.

Huitt, W. (2006). *Overview of classroom processes*. *Educational psychology interactive*. Valdosta, GA: Valdosta State University.

Ingle, L., Sleap, M., & Tolfrey, K. (2006). The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *Journal of Sports Sciences, 24*(9), 987-997.

Kelly, L.E., & Melograno, V.J. (2004). *Developing the Physical Education curriculum. An achievement-based approach*. Champaign, IL: Human Kinetics.

Lambert, L. (1996). Goals and outcomes. In S. Silverman & C. Ennis (Eds.), *Student learning in Physical Education: Applying research to enhance instruction* (pp. 149-169). Champaign, IL: Human Kinetics.

Mayorga-Vega, D., Merino-Marban, R., Vera-Estrada, F., & Viciana, J. (2014). Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology*, *46*(2), 227-233.

Mayorga-Vega, D., Viciana, J., & Cocca, A. (2013). Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics, 37*, 153-160.

Mayorga-Vega, D., Viciana, J., Cocca, A., & Rueda, B. (2012). Effect of a physical fitness program on physical self-concept and physical fitness elements in primary school students. *Perceptual and Motor Skills, 115*(3), 984-996.

McIlrath, D., & Huitt, W. (1995). *The teaching/learning process: A discussion of models*. Valdosta, GA: Valdosta State University.

Mesquita, I., Farias, C., & Hastie, P. (2012). The impact of a hybrid Sport Education-

Invasion Games Competence Model soccer unit on students' decision making, skill execution and overall game performance. *European Physical Education Review, 18*(2), 205-219.

Ministry of Education. (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria* [Royal Decree 1513/2006, the 7th December, laying down the Curriculum for Primary Education. In Spanish.]. Madrid: Government Gazette.

Newmann, F.M. (1992). *Student engagement and achievement in American secondary schools*. New York, NY: Teachers College Press.

Newmann, F.M., Marks, H.M., & Gamoran, A. (1995). Authentic Pedagogy: Standards that boost student performance. *Issues in Restructuring Schools, 8*, 1-12.

Pereira, J., Hastie, P., Araújo, R., Farias, C., Rolim, R., & Mesquita, I. (2015). A comparative study of students' track-and-field technical performance in sport education and in a direct instruction approach. *Journal of Sports Science and Medicine, 14*(1), 118-127.

Pérez-Pastur, R. (2010). Unidad didáctica sobre las capacidades físicas básicas [Teaching unit of fitness capacities. In Spanish.]. *Eduinnova*, *25*, 95-106.

Pesquie, P. *(1988).* L'Aprprentissage. [The learning. In French.]. *Revue EPS, 82,* 61-68.

Piéron, M. (1992). *Pédagogie des activités physiques et du sport.* [Pedagogy of physical activities and sport. In French.]. Paris: Revue EPS.

Robles, J., Giménez, F.J., & Abad, M.T. (2010). Motives for physical education teachers to select sport contents in secondary school. *Retos*, *18*(2), 5-8.

Román, B., Miranda, M.T., Martínez, M., Martínez, J.C., & Viciana, J. (2007). Transferencia del control de la ansiedad del patinaje en línea al esquí en educación física [Transfer of anxiety control from line-skate to skiing in Physical Education. In Spanish.]. *Revista Mexicana de Psicología, 20*, 392-393.

Salinas, F. (2011). *Conceptions of pre-service Physical Education teachers about planning.* (Unpublished doctoral dissertation. University of Granada). Granada.

Sánchez Rivas, E., Mayorga-Vega, D., Fernández Rodríguez, E., & Merino-Marban, R. (2014). Effect of a hamstring stretching programme during physical education lessons in primary education. *Journal of Sport and Health Research*, *6*(2), 159-168.

Senners, P. (2001). *La lección de Educación Física.* [The lesson in physical education. In Spanish.]. Barcelona: Inde.

Shortt, T., & Thayer, Y. (1999). Block scheduling can enhance school climate. *Educational Leadership, 56*(4), 76-81.

Siedentop, D. (1998). *Aprender a enseñar la Educación Física.* [Learning to teach physical education. In Spanish.]. Barcelona: Inde.

Siedentop, D., & Tanehill, D. (2000). *Developing teaching skills in physical education* (4th ed.). Mountain View, CA: Mayfield Publishing Company.

Silverman, S., Woods, A.M., & Subramaniam, P.R. (1999). Feedback and practice in physical education: Interrelationships with task structures and student skill level. *Journal of Human Movement Studies, 36*, 203-224.

Smith, W.W. (2011). Skills acquisition in physical education: A speculative perspective. *Quest, 63*(3), 265-274.

Stader, D.L. (2001). Block scheduling in small high schools: Perceptions from the field. *Rural Educator, 22*(3), 37-41.

UNESCO. (1962). *New methods and techniques in education: Report of a meeting of experts*. Paris, 12–20 March, 1962. Paris: UNESCO/ED/190.

UNESCO. (1974). Revised recommendation concerning technical and vocational education. *General Conference of UNESCO at its 18th session*, Paris, 19 November 1974. Retrieved in September of 2014 from https://www.unesco.org/education/pdf/23_29.pdf.

Van der Mars, H. (2006). Time and learning in physical education. In D. Kirk, D. Mcdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 191-213). London: Sage.

Viciana, J. (2002). *Planificar en Educación Física.* [Planning in Physical Education. In Spanish]. Barcelona: Inde.

Viciana, J., Lozano, L., Cocca, A., & Mayorga-Vega, D. (2012). Influence of the organizational system on motor engagement time in physical education on high school students. *Procedia-Social and Behavioral Sciences, 69*, 1160-1167.

Viciana, J., Salinas, F., & Cocca, A. (2008). Effects of an aerobic-capacity based program on cardiovascular risk factors in Spanish schools. *Circulation, 117*(11), e221.

Westerndorp, M., Houwen, S., Hartman, E., Mombarg, R., Smith, J., & Visscher, C. (2014). Effect of a ball skill intervention on children's ball skills and cognitive functions. *Medicine and Science in Sports and Exercise, 46*(2), 414-422.

Yáñez, J., & Castejón, F.J. (2011). La utilización de la transferencia para el aprendizaje de la táctica colectiva deportiva en Educación Secundartia [The use of transfer to learn collective tactics in sports in Secondary Education. In Spanish.]. *Infancia y Aprendizaje*, *34*(1), 95-10.

Zhang, T., Chen, A., Chen, S., Hong, D., Loflin, J., & Ennis, C. (2014). Constructing cardiovascular fitness knowledge in physical education. *European Physical Education Review, 20*(4), 425-443.

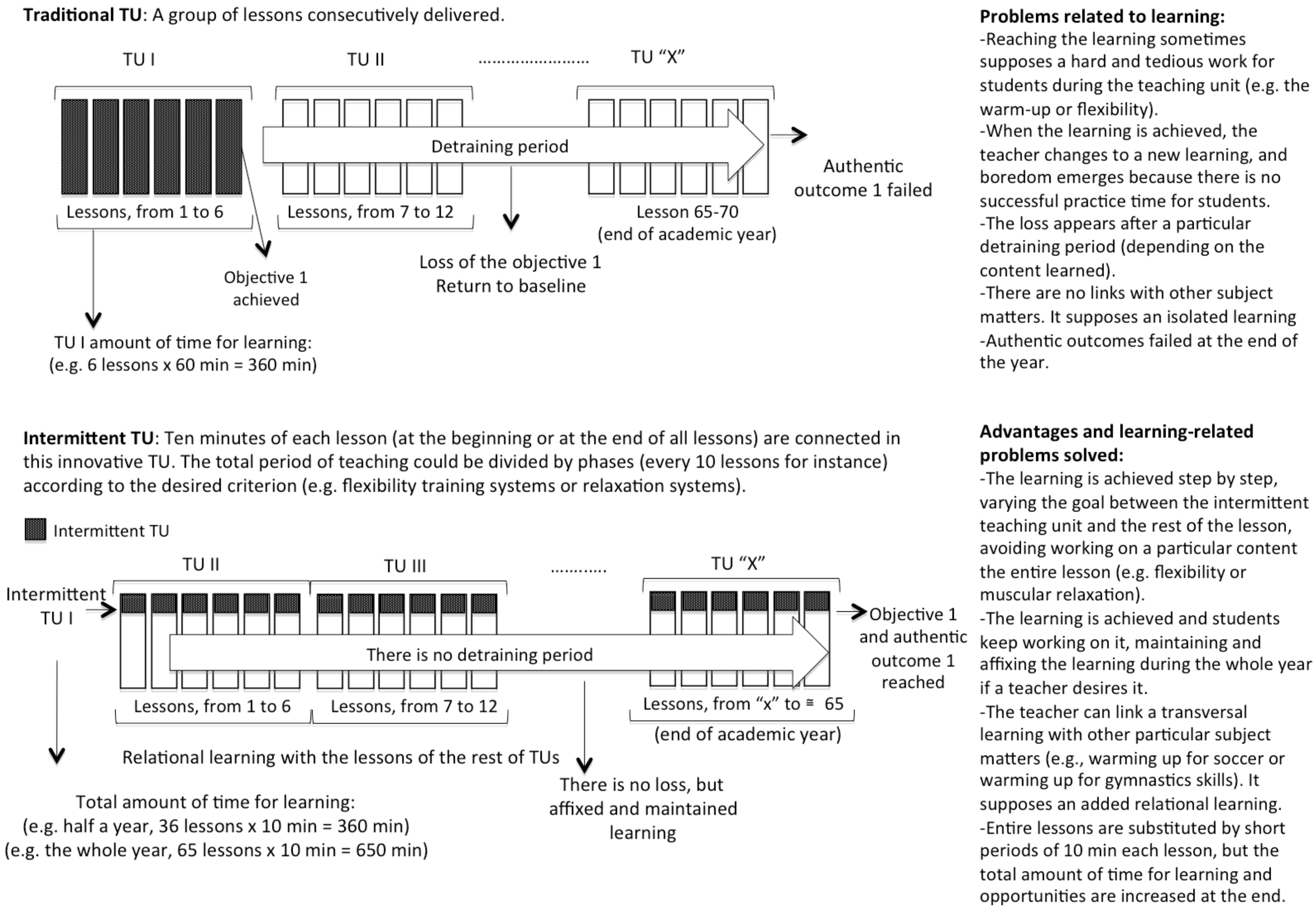


Figure 1. Comparison of intermittent and traditional teaching units.

Phase 1:
10 min from lessons 1 to 5

Example 1 (focused on students' autonomy during the warm-up): General warm-up directed by the teacher. Students reproduce and practice according the instructions of the teacher during the first 10 min of the lesson.

Example 2 (focused on the development of flexibility training methods): The Physical Education teacher develops dynamic systems of flexibility during the last 10 min of the lessons.

Example 3 (focused on the importance and awareness of alimentation): The Physical Education teacher explains alimentary guidelines (breakfast, lunch, dinner) and fosters the students' participation through a debate on their habits. Allocated at the first or last part of the lesson.

Phase 2:
10 min from lessons 6 to 15

Example 1 (focused on students' autonomy during the warm-up): General warm-up directed by the teacher, but using the students' proposals and variation of the tasks (during the first 10 min of the lesson).

Example 2 (focused on the development of flexibility training methods): The Physical Education teacher develops static systems of flexibility during the last 10 min of the lessons.

Example 3 (focused on the importance and awareness of alimentation): The Physical Education teacher explains the importance of hydration and fosters the students' participation through a debate on their habits. Allocated at the first or last part of the lesson.

Phase 3:
10 min from lessons 16 to 65

Example 1 (focused on students' autonomy during the warm-up): General warm-up directed by a couple of students each lesson (one explains the task and the other demonstrates it). During the first 10 min of the lesson.

Example 2 (focused on the development of flexibility training methods): The Physical Education teacher develops dynamic and static systems of flexibility according to the physical activity developed during the main part of the lesson (last 10 min of the lessons).

Example 3 (focused on the importance and awareness of alimentation):The Physical Education teacher explains the caloric balance and fosters the students' analysis of their daily habits, helping them to calculate their own caloric balance. Allocated at the first or last part of the lesson.

Figure 2. Examples of intermittent teaching units progression.

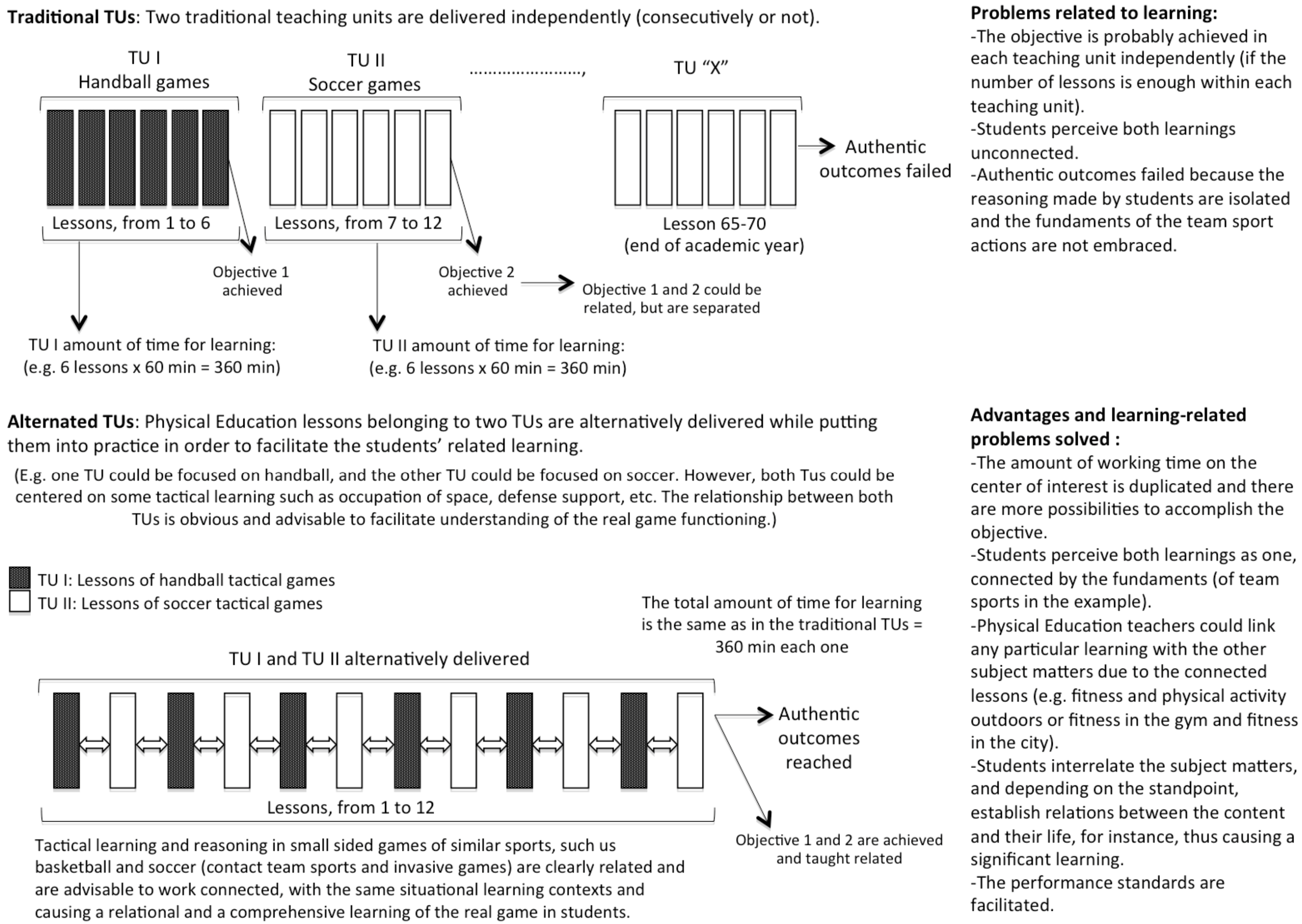


Figure 3. Comparison of alternated and traditional teaching units.

Alternated TU A

Tactical small sided games centered on soccer. The progression within the TU could be achieved by a varying number of players in each team, increasing the difficulty of tactical decision making.

Example 1 (focused on tactical learning in invasive team sports)

Alternated TU B

Tactical small sided games centered on basketball. The progression and the structure of the lessons are similar to the soccer TU, relating the learning of each lesson to the next one.

Alternated TU A

Fitness training methods developed indoors (applied to a gym or house). The progression within the TU could be achieved by different methods that students could use in the future and practice them.

Example 2 (focused on transferring the learning from the classroom to the students' life)

Alternated TU B

Fitness training methods developed outdoors (applied to open spaces such as parks, squares, sports fields, nature). The progression within the TU could be achieved by different methods that students could use in the future and practicing them.

Alternated TU A

Quantification of the students' exertion in order to classify energetic sources used in each activity realized. The progression within the TU could be achieved by different kinds of physical activities (i.e. orientation in nature, cycling, running, dancing, etc.). Perceived physical effort scales and manual methods to control the heart rate could also be used.

Example 3 (focused on the application of new technologies to physical activities)

Alternated TU B

Use of technologies to help students in quantification of their physical effort, and classifying the energetic sources used in each activity realized. The progression within the TU could be achieved by different kinds of technologies used in each physical activity proposed (digital compass, i-cycling or sport runner apps for smartphones and IPhones, zumba fitness videos, etc.). Heart rate monitors could be used to control exercise intensity.

Figure 4. Examples of alternated teaching units.

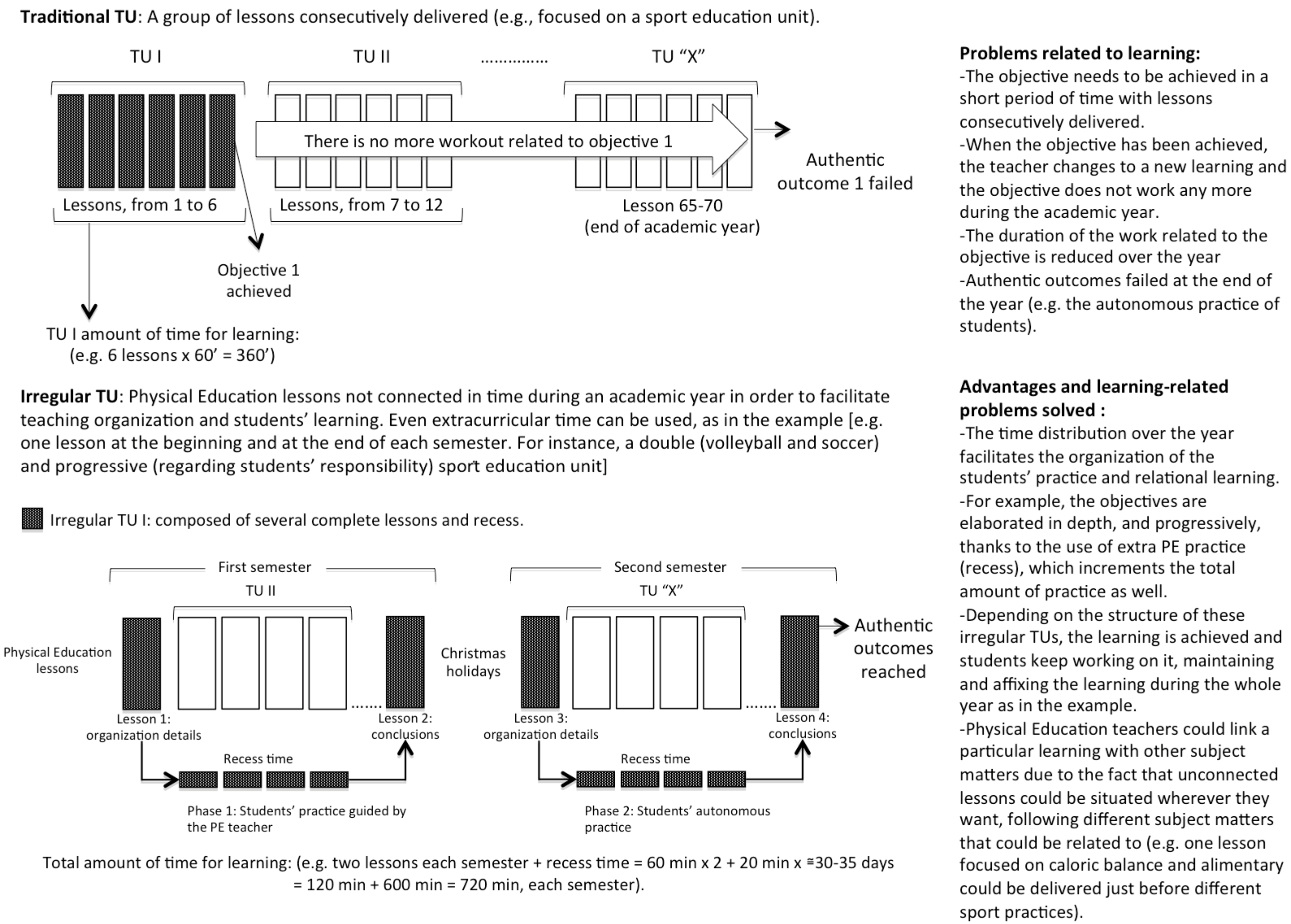


Figure 5. Comparison of irregular and traditional teaching units.

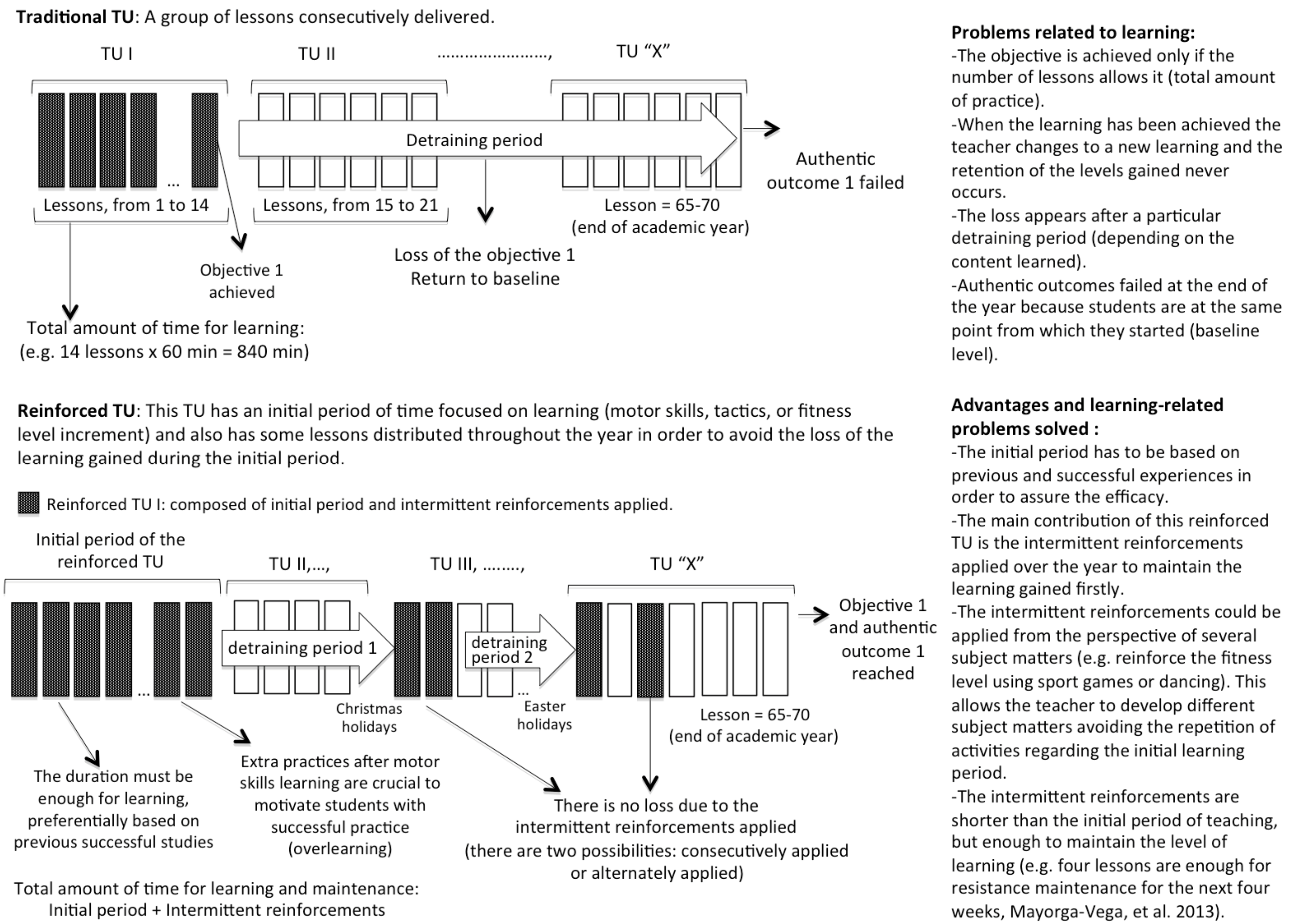


Figure 6. Comparison of reinforced and traditional teaching units.

Paper IV

MODELO DE APRENDIZAJE EXITOSO EN EDUCACIÓN FÍSICA Y SU MANTENIMIENTO. ESTUDIO DEL EFECTO DEL REFUERZO INTERMITENTE SOBRE LA CONDICIÓN FÍSICA

[SUCCESSFUL LEARNING MODEL IN PHYSICAL EDUCATION AND ITS MAINTENANCE. THE INTERMITTENT REINFORCEMENT APPLIED TO PHYSICAL FITNESS]

Viciana, J., Mayorga-Vega, D., & Cocca, A.

Revista Iberoamericana de Psicología del Ejercicio y el Deporte

2014, *9*(1), 155-171.

ISSN 1886-8576

Revista Iberoamericana de Psicología del Ejercicio y el Deporte
Vol. 9, nº 1 pp. 155-171

MODELO DE APRENDIZAJE EXITOSO EN EDUCACIÓN FÍSICA Y SU MANTENIMIENTO. ESTUDIO DEL EFECTO DEL REFUERZO INTERMITENTE SOBRE LA CONDICIÓN FÍSICA

Jesús Viciana Ramírez, Daniel Mayorga-Vega y Armando Cocca
Facultad de Ciencias de la Actividad Física y el Deporte.
Universidad de Granada. España

RESUMEN: Conseguir aprendizajes exitosos en Educación Física es una tarea compleja que parte desde la propia planificación. Existen evidencias científicas de falta de planificación entre los docentes en ejercicio, así como dudas a la hora de realizarla que agravan este problema. El propósito de este artículo es describir el proceso de la planificación de aprendizajes motores y de desarrollo de la condición física en el contexto escolar, analizando las circunstancias que rodean a este proceso, con el fin de orientar a los profesores a planificar y que garanticen cierta eficacia en la intervención. Finalmente se expone el resumen de una investigación con la aplicación de un tipo de refuerzo intermitente para mantener los logros conseguidos en una intervención previa como ejemplo de lo explicado anteriormente.

PALABRAS CLAVE: Aprendizaje motor; olvido; retención; condición física-salud.

Manuscrito recibido: 24/11/2012
Manuscrito aceptado: 13/11/2013

Dirección de contacto: Jesús Viciana Ramírez. Facultad de Ciencias de la Actividad Física y el Deporte, Universidad de Granada. Carretera de Alfacar s/n, 18011, Granada, España.
Correo-e.: jviciana@ugr.es

SUCCESSFUL LEARNING MODEL IN PHYSICAL EDUCATION AND ITS MAINTENANCE. THE INTERMITTENT REINFORCEMENT APPLIED TO PHYSICAL FITNESS

ABSTRACT: Achieving a high level of motor learning and physical fitness in the Physical Education setting is a complex task for teachers, and it starts from the planning. Some evidence shows a lack of planning from in-service teachers as well as many doubts in the planning process. The aim of this article is to describe the process of planning motor learning and physical fitness in a physical education setting. As well as analyzing the circumstances around this process, in order to guide teachers in planning, guaranteeing future success in interventions. Finally, we expose the summary of research with the application of an intermittent reinforcement to maintain the gains achieved in a previous intervention as an example.

KEYWORDS: Motor Learning; forget; retention; health-related physical fitness.

MODELO DE APRENDIZAGEM BEM SUCEDIDO EM EDUCAÇÃO FÍSICA E SUA MANUTENÇÃO. ESTUDO DO EFEITO DO REFORÇO INTERMITENTE SOBRE A CONDIÇÃO FÍSICA

RESUMO: Conseguir aprendizagens bem-sucedidos na Educação Física é uma tarefa complexa que começa desde a própria planificação. Existem evidencias científicas da falta de planificação entre os docentes em exercício, assim como duvidas à hora de realiza-la, que agravam este problema. O propósito deste artículo é descrever o processo da planificação de aprendizagens motores e de desenvolvimento da condição física no contexto escolar, analisando as circunstâncias que cercam este processo, com o fim de orientar os professores para planificar, e que garantem certa eficácia na intervenção. Finalmente expõe-se o resumo duma investigação com a aplicação dum tipo de reforço intermitente para manter os logros conseguidos na intervenção prévia como exemplo do explicado anteriormente.

PALAVRAS-CHAVE: Aprendizagem motor; esquecimento; retenção; condição física-saúde

La evidencia científica nos pone alerta ante los principales problemas que rodean a la planificación de Educación Física (EF) en la mayor parte de los países: a) la falta de planificación reflexiva y que contemple toda una etapa; b) la escasa supervisión de las funciones docentes por parte de las administraciones públicas (Hardman, 2008); c) y la falta de planificación escrita, unida al uso de la improvisación (Viciana y Zabala, 2004).

Los docentes de EF y de cualquier materia deben planificar su quehacer docente de manera obligatoria desde las directrices del currículo nacional, en periodos de tiempo que normalmente son años o cursos académicos de aproximadamente nueve meses. Además, deben contemplar los periodos vacacionales, semanas culturales y días festivos en dichas programaciones. En este tiempo, los docentes deben encajar periodos de trabajo de diferente extensión para conseguir los objetivos educativos hacia los alumnos.

Por tanto, una programación anual se concreta en periodos más pequeños de enseñanza, incluso la mayoría de las veces desconectados entre sí o independientes unos de otros. De todos es sabido que la distribución de las programaciones anuales en EF se ha constituido desde hace décadas en subconjuntos temporales denominados 'unidades didácticas' (UD) o 'unidades de enseñanza' (Piéron, 1992).

Aunque la UD es un elemento de mayor concreción que la programación de aula, y por tanto forma parte de ella, es de gran importancia porque representa la estructura principal de la planificación. Los profesores de EF por norma general cuando planifican, piensan en unidades didácticas y no en programas anuales de intervención, para así componer y planificar su trabajo diario. Por ello, entre otras causas, la UD permite una gran variabilidad en su diseño y dota al profesor de un carácter personal y particular, distinguiéndolo de los demás.

Por otra parte, el elemento principal de la UD y de cualquier programación es el objetivo educativo. Un objetivo educativo debe estar compuesto por un verbo que indica el tipo de aprendizaje que se espera que el alumno consiga en el futuro. Habrá ocasiones que usando verbos que indiquen actividad cognitiva (p.ej. saber, conocer, aplicar o relacionar), el objetivo se encaminará a mejorar los conceptos, sus relaciones, su aplicación y en general, a mejorar la parte cognitiva del alumnado. En otras ocasiones el verbo indicará una acción motriz, procedimientos y destrezas motrices a mejorar en el futuro (p. ej. realizar, desarrollar, aplicar motrizmente, mejorar una habilidad, vivenciar, o experimentar). Y finalmente otros verbos indicarán una mejora esperada en las actitudes de los alumnos (p. ej. tomar conciencia de, respetar, valorar o considerar a los demás). Además, el objetivo se compone también de un contenido sobre el cual recae la acción del verbo empleado. Los contenidos en EF son muy variados y en el ámbito escolar generalmente suelen encuadrarse en juegos, deportes, condición física, salud, expresión corporal, capacidades coordinativas, habilidades y actividades en el medio natural, fundamentalmente. Estos contenidos además se pueden especificar más o menos, que conjuntamente con el empleo de un verbo u otro, hará del objetivo una 'empresa' más o menos difícil de conseguir.

El objetivo marca la intervención que el profesor necesita aplicar para conseguirlo tras un periodo de interacción con el alumno, marca las intensidades de las tareas, la duración de las mismas, los estilos de enseñanza empleados, la comunicación del profesor con el alumno y de los alumnos entre sí, así como incluso las normas sociales que rigen los comportamientos en clase.

Por todo lo dicho anteriormente, podemos concluir que planificar la EF escolar no es tarea fácil. En resumidas cuentas, una planificación se va a caracterizar y condicionar por el objetivo marcado y, según él, debemos configurar un periodo de trabajo suficiente y necesario para conseguirlo, así como diseñar y proponer una intervención adecuada.

Planteamiento del problema

La programación anual de EF ha sido y sigue siendo considerada en muchos casos, segmentos de tiempo más pequeños, a veces con la misma duración, y rellenas de contenidos elegidos por criterios muy diversos, en ocasiones puramente por preferencias personales o por tendencias provocadas por el material disponible. Sin embargo, el criterio más efectivo para tomar estas dos decisiones tan importantes en la planificación (la duración y los contenidos elegidos) debe ser el propio objetivo, su naturaleza y su complejidad. Delaunay y Pineau (1989) resaltaban que "muy a menudo, la unidad didáctica, centrada en la noción de tiempo (tantas horas, tantas semanas...), olvida la importancia de las nociones de objetivos y contenidos. Se considera como un recipiente, una forma que rellenar... cuando el tiempo debería ser elegido en función de los objetivos y contenidos determinados" (Seners, 2001, p. 83).

De entre los posibles objetivos que la EF puede proponerse con sus alumnos en sus UD, destacamos como posibilidades los ocho grupos que nombraba Viciana (2002): (1) adquirir conductas motrices nuevas (habilidades, destrezas, patrones motores, etc.); (2) modificar conductas motrices ya existentes (perfeccionar gestos ya aprendidos anteriormente); (3) desarrollar la condición física del alumno; (4) vivenciar métodos correctos de práctica física (busca que el alumno experimente la vivencia de un método para que lo aprenda y lo use en el futuro, pero no necesariamente que desarrolle lo que el método persigue); (5) recrearse y disfrutar personalmente con la actividad física (incluye aquí todo lo relativo a la satisfacción y motivación intrínseca y extrínseca que produce la actividad física y los factores que la rodean); (6) adquirir hábitos positivos de práctica física (adherencia a la práctica y cualquier otro hábito como pueden ser los de carácter higiénico o ergonómicos); (7) asimilar conceptos relacionados con la EF; y (8) adquirir actitudes y valores positivos con el entorno y con las personas.

Si analizamos estos grupos, vemos que los objetivos relativos a la recreación y motivación, así como los relativos a la asimilación de conceptos se pueden conseguir con intervenciones cortas, de escasas sesiones. Sin embargo, la adquisición y perfeccionamiento de acciones motrices (aprendizaje de patrones motores y habilidades específicas) y el aumento de la condición física en los alumnos, precisan de intervenciones más largas y con un diseño especial de las tareas. Finalmente, los grupos de objetivos encaminados a la adquisición de hábitos de práctica (adherencia a la actividad física) así como el cambio en las actitudes (sobre todo si hablamos de comportamientos y no de predisposición) pensamos que son producto de una intervención más prolongada en el tiempo y que deben trabajarse durante años, consiguiéndose poco a poco.

Aunque la práctica más normal entre los docentes de EF es planificar las UD combinando estos grupos de objetivos (p. ej. que el alumno aprenda conceptos nuevos, practique métodos correctos de práctica física y disfrute con las propias tareas propuestas), la primera reflexión que debemos hacer es *analizar si entre nuestros objetivos se encuentra alguno de los que precisan intervenciones más largas o simplemente con unas pocas sesiones son suficientes para conseguirlos*. A veces estos objetivos más pretenciosos son exigidos desde las recomendaciones de los propios currículos nacionales de un país determinado, como es el caso de la Ley Orgánica de Educación española (Real Decreto 1531/2006 de Enseñanzas Mínimas para Secundaria, 2007) en la etapa de Secundaria, donde se persigue por ejemplo el incremento de la condición física del alumno, principalmente centrada en las cualidades de resistencia, fuerza y flexibilidad como cualidades más representativas de la salud.

La siguiente cuestión lógica que un docente debe plantearse es *¿Qué duración debe tener la UD para asegurar cierta eficacia en la consecución de estos aprendizajes u objetivos más costosos?* Hébrard (1986) comenta que es difícilmente aceptable hablar de aprendizaje (susceptible de permanecer en el olvido) gracias a una decena de sesiones de práctica. Viciana (1998) habló de un número de sesiones entre 15 y 20, cuestión comprobada empíricamente en el aprendizaje de habilidades específicas de voleibol con 18 sesiones en un centro de secundaria, obteniendo resultados positivos. Viciana, Salinas y Cocca (2008) también consiguieron incrementar la resistencia aeróbica de un grupo de alumnos de secundaria con 18 sesiones. Piéron (1988) comenta que una UD tiene normalmente una duración de 6 a 8 semanas, o sea, de 12 a 16 sesiones. Kelly y Melograno (2004) por ejemplo, concretan que un número 'normal' de sesiones para UD típicas es de cuatro a 12, o sea, de dos a seis semanas. Seners (2001) –también señala que las directrices administrativas francesas fijan entre 10 y 12 sesiones de práctica la cantidad de sesiones mínimas para el aprendizaje, diferenciando entre institutos y colegios, y comenta que una cantidad de 6 a 12 son significativas, aunque reconoce la dificultad de fijar una regla exacta.

Realmente, todos los autores coinciden en la dificultad de fijar un número de horas o sesiones para asegurar el aprendizaje. Variables como el tipo de alumnos, las contingencias ambientales, el contenido, la intervención del profesor en el aula, etc., serán determinantes en el aprendizaje y en el establecimiento de un tiempo mínimo necesario para su consecución. Sólo nos resta concluir a este respecto que la cantidad mínima de sesiones para conseguir aprender una habilidad o aumentar la condición física del alumno dependerá de muchos factores y sólo con la innovación docente (implementando programas fundamentados) podremos asegurar su consecución. Así, aquellos docentes sin experiencia, deberán consultar las experiencias de otros para proponer justificadamente una duración del trabajo con ciertas garantías de éxito.

Posteriormente, se da una segunda incógnita igualmente importante para el aprendizaje, *¿cómo podemos retener los aprendizajes realizados?* Surge ahora el problema de mantener la condición física o los aprendizajes conseguidos en los alumnos para no caer en el olvido o volver a los valores iniciales. A este respecto Seners (2001) señala que 12 meses después de una unidad didáctica de cinco lecciones en una misma actividad, el alumno estará de nuevo en el nivel más bajo de práctica. A menudo, nos olvidamos de reforzar los logros conseguidos para evitar este descenso y nos conformamos con haber conseguido un nivel aceptable tras la aplicación de un programa de intervención. Conseguir determinados logros no tiene mucho sentido si después no los conservamos, sobre todo cuando hablamos de incrementos de la condición física, por tratarse de un importante marcador de salud entre los jóvenes (Ortega, Ruiz, Castillo y Sjöström, 2008), entre otros motivos [como por ejemplo que la condición física incrementa o mantiene el autoconcepto físico que actúa de mediador también en la práctica física (Mayorga-Vega et al., 2012)].

Además, se une a este problema el hecho de que los profesores normalmente deben desarrollar un gran número de contenidos de EF a lo largo de su programación, lo cual provoca dos circunstancias: (a) el profesor programa pocas sesiones para cada contenido, lo cual no permite que el alumno consiga el objetivo de aprender o desarrollar la condición física; (b) a menudo el profesor, presa del tiempo, cambia la situación o la actividad desde el momento en que el alumno tiene éxito o consigue un nivel aceptable. Tras una intervención en la que el alumno está continuamente esforzándose para aprender o mejorar (cometiendo errores y reconduciendo su propio aprendizaje), cuando apenas lo consigue, le cambiamos el tipo de aprendizaje introduciendo una nueva habilidad que aprender. Esto provoca en el alumnado una desmotivación que no favorece el clima de aprendizaje, tendiendo al fracaso y al abandono (Viciana, 2002).

Este procedimiento es poco motivador ya que tiene como consecuencia enfrentar siempre al alumno con una tarea en la que fracasa, en lugar de permitirle seguir en situación de éxito. Aparece entonces el concepto de “sobreaprendizaje” (Seners, 2001), como la cantidad de práctica “extra” para fijar los aprendizajes en los alumnos, a lo que añadimos '... y permitirle practicar con éxito tras haber aprendido'. También ha sido llamado "resistencia a la extinción" por Le Ny (1980), quien comenta que una reacción es más difícil de apagar si ha sido reforzada que si lo ha sido relativamente poco.

Sin embargo, otros autores comentan el entorpecimiento de lo que se aprende cuando la cantidad de práctica es excesiva. Pesquie (1988) señala que cuando un aprendizaje es largo, la repetición plantea algunos problemas y la repetición puede desarrollar un estado de fatiga que poco a poco puede volver al organismo refractario a una nueva reacción. Así pues, es necesario emplear repeticiones agrupadas en series y series separadas por intervalos de tiempo que permitan la recuperación. Como comenta Seners (2001), esto lo podemos aplicar a las repeticiones de unas tareas durante la sesión, o las sesiones de una UD o incluso las UD de un curso académico.

De aquí deducimos que una solución válida es la repetición de UD en torno a un mismo núcleo de interés para provocar el aprendizaje y su retención posterior, con un intervalo de tiempo entre ellas. Seners (2001) nos comenta que es necesario algún tiempo para asimilar bien lo que se ha aprendido. Para evitar una masificación durante el año, se pueden realizar dos unidades didácticas de una misma actividad en lugar de una sola de larga duración. Separadas algunos meses, permiten la maduración durante el periodo interunidades didácticas (desentrenamiento). Además, evitan dejar al alumno sin práctica de una misma actividad durante un periodo de tiempo largo. Así, la misma actividad se aborda dos o más veces al año, aplicando lo que llamamos 'refuerzos intermitentes' (Le Ny, 1980) o reactivación del aprendizaje (López, 1992).

El modelo de aprendizaje exitoso en Educación Física

En el modelo de planificación exitosa de aprendizajes motores e incremento de la condición física, representado en la figura 1, observamos todos los factores descritos anteriormente. Vamos a describir por fases los pasos a seguir en este modelo.

1) En primer lugar el profesor, tras proponerse como objetivo el aprendizaje de una habilidad técnica o táctica en un deporte, o el incremento de condición física en una UD, debe determinar el tiempo necesario para aprender o mejorar esa habilidad. Para determinar esa cantidad de sesiones de la UD en cuestión, debemos consultar si autores anteriores han llevado a cabo algún aprendizaje similar en contextos parecidos y adoptar su temporalización así como su metodología, para garantizar cierta eficacia en el aprendizaje. Esta revisión de estudios anteriores cobra más importancia cuando el profesor no tiene experiencias previas en su propio contexto. En caso de que no se encuentren experiencias y recomendaciones de otros autores sólo nos queda la propia innovación en la práctica, creando nuestro propio cuerpo de conocimientos y construyendo autónomamente el currículo para crear los primeros antecedentes de ese aprendizaje.

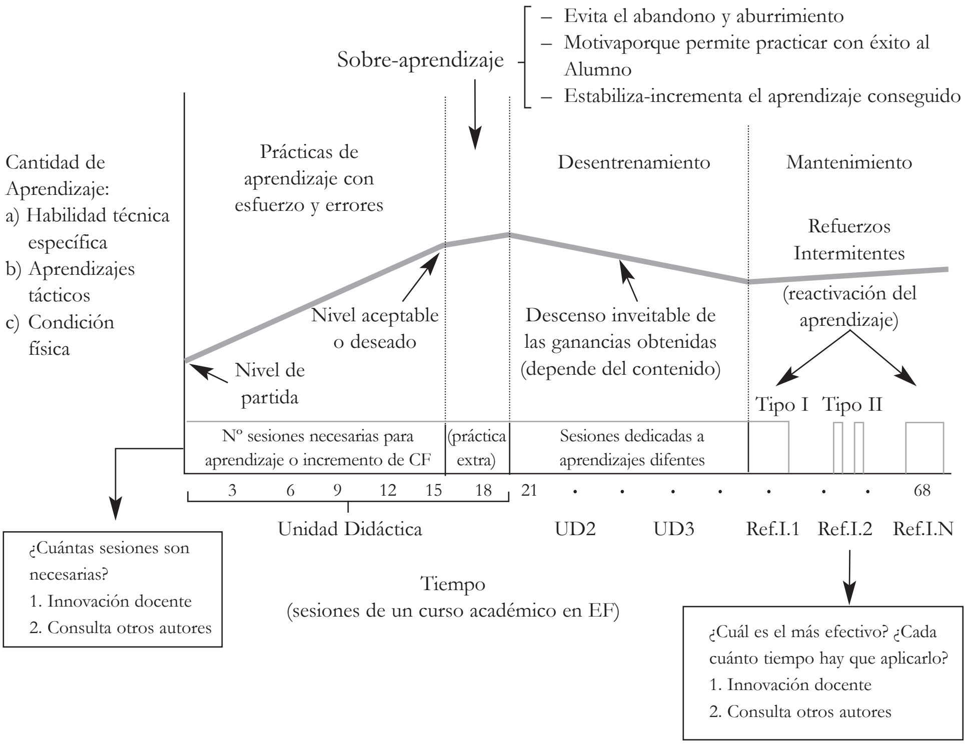


Figura 1. Modelo de planificación exitosa de aprendizaje motor e incremento de la condición física y su mantenimiento en EF.

2) En segundo lugar, tras la aplicación del primer programa de aprendizaje, debemos tener en cuenta que la práctica del alumno se extienda lo necesario para tener unos días de práctica con éxito tras aprender, aplicando el sobre-aprendizaje y garantizando la motivación y la consolidación del aprendizaje del alumno.
3) Debemos consultar igualmente en la literatura, el tiempo que tarda el alumno en perder los aprendizajes conseguidos, para no caer a los valores iniciales por un desentrenamiento demasiado largo. Y consecuentemente aplicar un refuerzo intermitente, consistente en nuevas prácticas con metodologías similares o iguales y con una extensión menor que el programa inicial.
4) Finalmente, haremos esto cada vez que se precise con la finalidad de acabar el año académico con un nivel aceptable. El tiempo que no tenemos al alumno

en clases, como las vacaciones de verano, debemos dotar de autonomía suficiente al alumno (sobre todo en cursos de secundaria) para que los estímulos de aprendizaje o de incremento de condición física puedan darse de manera autónoma.

El tipo de refuerzo intermitente, bien de manera discontinua (alternando sesiones de estimulación y sesiones de otros contenidos), o bien de manera continua con sesiones sucesivas de estímulo del aprendizaje inicial, deben ser aplicados y evaluados para comprobar su efectividad en cada caso. Existe actualmente una necesidad de mayor experimentación de este modelo en EF.

A continuación, vamos a exponer resumidamente los logros alcanzados en una intervención sobre la condición física y su mantenimiento en clases de EF con alumnos de último curso de primaria en España.

Efecto del entrenamiento de la condición física-salud en alumnos de primaria españoles y su mantenimiento con un refuerzo intermitente discontinuo

El Ministerio de Educación Español propone como contenido (dentro del bloque de Actividad física y salud) la 'Mejora de la condición física del alumno' y como objetivo para tercer ciclo de primaria (alumnos de 10-12 años) en su criterio de evaluación número seis 'Mostrar conductas activas para incrementar globalmente la condición física, ajustando su actuación al conocimiento de las propias posibilidades y limitaciones corporales y de movimiento' (Real Decreto 1513/2006 de las Enseñanzas Mínimas para Primaria en la LOE, 2006:43080). Por ello, nos propusimos aplicar un programa a corto plazo para incrementar la condición física de los alumnos de tercer ciclo de Primaria de un colegio de Granada, España.

MÉTODO

Participantes

Participaron un total de 72 alumnos (40 niños y 32 niñas, edad media = 11.1 años ± 0.38 años) de cuatro grupos de clase diferentes que dividimos en dos grupos balanceando el género: experimental (n = 35) y control (n = 37). El 75% de ellos participaban de manera esporádica durante la semana en programas extraescolares deportivos, y se les instó a continuar con su actividad diaria sin modificaciones. Sin embargo, controlamos que no hubieran actividades encaminadas al desarrollo de la condición física.

Diseño

En la presente investigación se usó un diseño cuasi-experimental *cluster randomized controlled trial* con tres medidas repetidas (pretest, postest, retest). Lo observamos en la figura 2.

Grupo Experimental	O_1———-X_1———-O_2	Desentrenamiento	X_2 O_3
Grupo Control	O_1——————-O_2	Desentrenamiento	— O_3

Figura 2. Diseño de la investigación.

Nota: Donde O_1 y O_2 son las medidas pre y post-tratamiento; X_1 es el programa aplicado para incrementar la Condición Física; X_2 es el Refuerzo Intermitente aplicado de manera discontinua; y O_3 es la medida retest.

Instrumentos

Se evaluó a todos los alumnos para medir su nivel de condición física con las siguientes pruebas pertenecientes a la batería Eurofit: resistencia cardiovascular (Course Navette), fuerza abdominal (Abdominales de 30 s), y fuerza de brazos (Suspensión con flexión de brazos). Se tomaron estas pruebas por ser estandarizadas, validadas internacionalmente (Consejo de Europa Comité para el Desarrollo del Deporte, 1988) y aplicadas en numerosos casos en contextos educativos (Ardoy et al., 2011; Dorgo et al., 2009).

Procedimiento

Siguiendo las recomendaciones de autores anteriores (Seners, 2001; Viciana, 2002) propusimos un programa de dos veces por semana durante ocho semanas. En el programa se completó un total de 14 sesiones (más dos clases que coincidieron con días festivos). Igualmente, seguimos las sugerencias de otros autores para la metodología, usando el circuito como organización de las tareas, ya que aumentan considerablemente el tiempo de compromiso motor (Lozano et al., 2009), y usando tiempos de trabajo y descanso como método de trabajo en las estaciones o tareas del circuito. El circuito consistía en ocho estaciones de fuerza-resistencia que mantenían una intensidad de trabajo alta y se trabajaba la fuerza de piernas, brazos y tronco. Este circuito se repetía dos veces en cada sesión y al final de cada uno se realizaba un juego de carreras continuo y global con todos los alumnos durante cinco minutos. Concretamente la progresión de los tiempos de trabajo del circuito se realizó durante las 14 sesiones desde los 15 a los 35 segundos de trabajo por estación y se disminuyó el descanso desde los 45 a los 25 segundos. En este tiempo de trabajo el alumno debía realizar el ejercicio el mayor número de veces posible sin desvirtuar el movimiento de dichas tareas (Faigenbaum et al., 2002; Faigenbaum et al., 2005).

Al comienzo y al final del programa se evaluaron a todos los alumnos con las pruebas descritas anteriormente para comprobar su eficacia. Después del tratamiento, los alumnos estuvieron cuatro semanas inactivos por las vacaciones de Navidad y en la quinta semana se comenzó a aplicar un refuerzo intermitente que consistió en cuatro sesiones aplicadas de manera discontinua (intercalando sesiones de otros contenidos) durante cuatro semanas más. En estas sesiones se volvió a realizar la misma metodología que en las sesiones del programa inicial. Finalmente se midieron a todos los alumnos con las pruebas seleccionadas de la batería Eurofit para comprobar el grado de mantenimiento de los logros iniciales. El grupo control realizó clases de juegos tradicionales e introducción a los deportes de voleibol y baloncesto durante las sesiones del tratamiento inicial y mantenimiento, con lo cual no realizaron los circuitos de fuerza-resistencia, sino las clases 'normales' de EF previstas por el profesor en su programación anual.

Análisis de datos

Se aplicó un análisis de varianza (ANOVA) de dos factores con el grupo (experimental y control) como factor inter-sujetos y el tiempo (pretest, postest, y retest) como factor de medidas repetidas sobre las variables dependientes estudiadas. Posteriormente, en el análisis *post-hoc* para evaluar el factor intra-grupo se aplicó la corrección de Bonferroni. Como la variable suspensión con flexión de brazos no seguía una distribución normal, previamente se transformaron los datos logarítmicamente (Bland y Altman, 1996). El análisis estadístico se realizó mediante el paquete estadístico SPSS 15.0 para Windows (SPSS® Inc., Chicago, IL). El nivel de significación estadística se estableció en $p < .05$.

RESULTADOS

Los resultados fueron los siguientes: (i) para la fuerza abdominal [$F(2. 63) = 4.636$; $p < .05$]. El posterior análisis *post-hoc* de Bonferroni mostró el significativo incremento del pretest al postest ($p < .05$) y del pretest al retest ($p < .01$) para el grupo experimental, mientras que para el grupo control no hubo diferencias ($p > .05$); (ii) para la fuerza de brazos [$F(2.63) = 5.994$; $p < .01$]. El *post-hoc* para el grupo experimental de pretest a posttest ($p < .01$) y de pretest a retest ($p < .001$), e incluso hubo indicios de significación entre posttest y retest ($p < .10$). Para el grupo control no hubo diferencias (p > 0,05); (iii) para la resistencia cardiovascular [$F(2, 64) = 5.230$; $p < .01$]. El ANOVA con la penalización de Bonferroni mostró que el grupo experimental incrementó del pretest al posttest ($p < .05$), y aunque desde el pretest al retest sólo hubo indicios de significación ($p < .10$), no se encontraron diferencias del posttest al retest ($p > .05$), demostrando la validez del refuerzo intermitente aplicado. Igualmente no hubo diferencias estadísticamente significativas para el grupo control ($p > .05$).

En las figuras 3-5 se observan los datos descriptivos de las tres mediciones en las tres variables medidas.

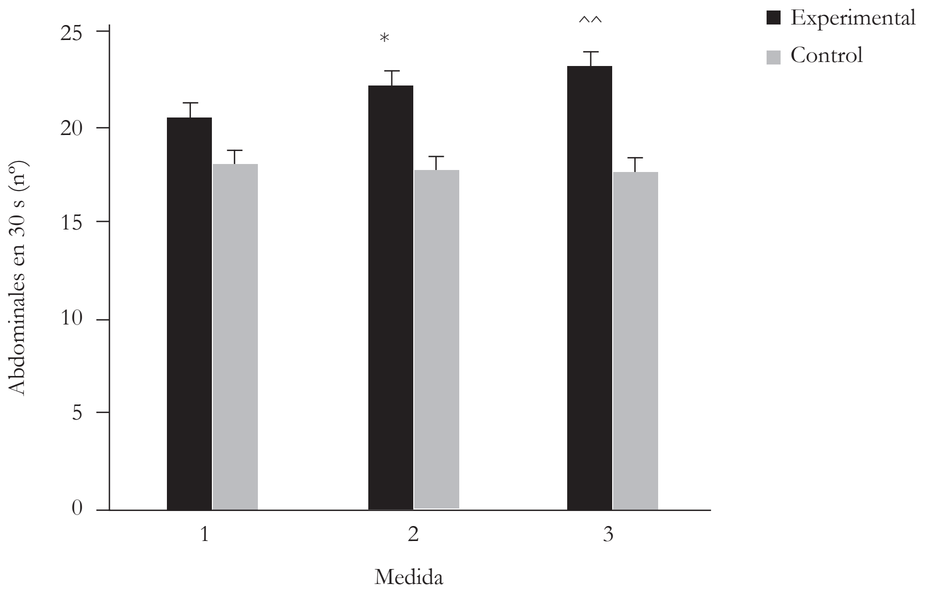


Figura 3. Abdominales en 30 segundos en el grupo experimental y grupo control en las tres medidas.

Nota: Valores medios y la barra de error representa el error típico. El análisis de la varianza mostro efecto de interacción grupo-tiempo (p < .05). Análisis post-hoc de Bonferroni: Cambios estadísticamente significativos desde la medida 1 a la 2 (*p < .05) y de la 1 a la 3 (^^p < .01).

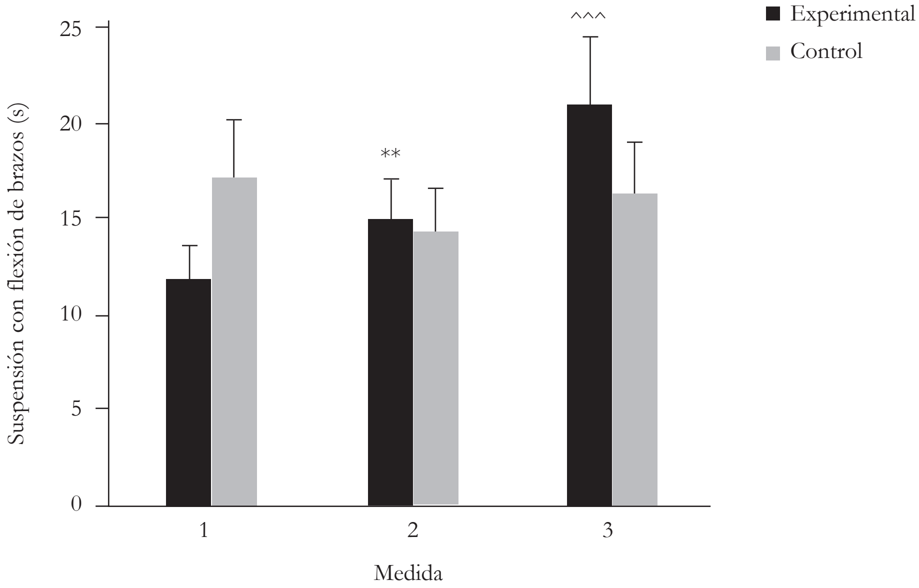


Figura 4. Suspensión de brazos en el grupo experimental y grupo control en las tres medidas.

Nota: Valores medios y la barra de error representa el error típico. El análisis de la varianza mostro efecto de interacción grupo-tiempo ($p < .01$). Análisis post-hoc de Bonferroni: Cambios estadísticamente significativos desde la medida 1 a la 2 (**$p < .01$) y de la 1 a la 3 (^^^$p < .001$).

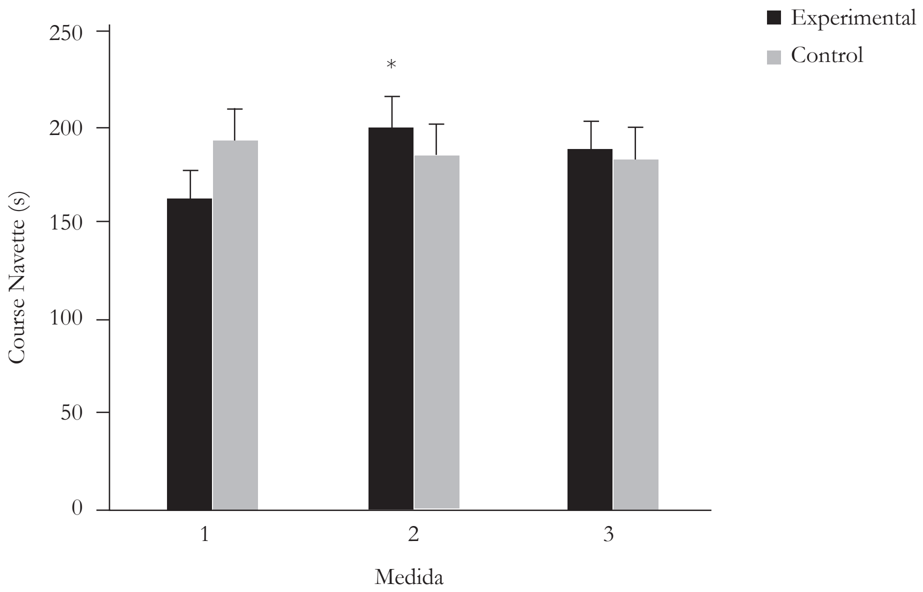


Figura 5. Course Navette en el grupo experimental y grupo control en las tres medidas.

Nota: Valores medios y la barra de error representa el error típico. El análisis de la varianza mostro efecto de interacción grupo-tiempo ($p < .01$). Análisis post-hoc de Bonferroni: Cambios estadísticamente significativos desde la medida 1 a la 2 (*$p < .05$).

CONCLUSIÓN

La planificación de la EF es una tarea compleja, siendo esto asumido en otros campos como el entrenamiento deportivo o incluso la gestión, pero a veces olvidado en el campo educativo. Esta complejidad hace que debamos analizar los objetivos que nos proponemos y obrar en consecuencia con programas suficientemente extensos en el tiempo (UD con suficiente práctica para lograr los aprendizajes).

Proponemos el modelo de aprendizaje exitoso en EF para planificar con ciertas garantías de éxito un programa de intervención en clase y lograr aprendizajes técnicos (Viciana, 1998), tácticos (Mesquita, 2006; Mesquita et al., 2005), o el incremento de la condición física (Viciana, Salinas y Cocca, 2008). Hemos descrito las fases y hemos representado el modelo explicándolo para su comprensión y realización en el futuro por los docentes de EF que estimen oportuno aplicarlo.

En nuestro caso de aplicación de dicho modelo en el último epígrafe, hemos querido poner un ejemplo explicando las fases del modelo y mostrando la fundamentación o los porqués de nuestra intervención, apoyándonos en autores anteriores y garantizando así la eficacia de la intervención.

En primer lugar, demostramos que es posible incrementar la condición física en clases de EF con un programa a corto plazo, desmitificando las creencias que existen al respecto sobre la imposibilidad de conseguirlo con solo dos sesiones semanales. Y en segundo lugar, demostramos la validez de un refuerzo intermitente para mantener los logros conseguidos con el primer programa (DeRenne et al., 1996), aplicado de manera discontinua durante cuatro semanas y antes de perder las ganancias obtenidas (Da Fontoura et al., 2004; Ingle et al., 2006; Tsolakis et al., 2004). Finalmente, instamos a los docentes de EF a probar el modelo con aprendizajes de diferente naturaleza, como son los aprendizajes motores de habilidades técnicas o tácticas de deportes, e igualmente a probar diferentes tipos de refuerzos intermitentes y comprobar su eficacia en el contexto de la EF escolar.

REFERENCIAS

Ardoy, D. N., Fernández-Rodríguez, J. M., Ruiz, J. R., Chillón, P., España-Romero, V., Castillo, M. J., y Ortega, F. B. (2011). Mejora de la condición física en adolescentes a través de un programa de intervención educative: Estudio EDUFIT. *Revista Española de Cardiología*, *64*, 484-491.

Bland, J. M., y Altman, D. (1996). Transforming data. *British Medical Journal*, *312*(7033), 770.

Consejo de Europa Comité para el Desarrollo del Deporte (1988). *Eurofit: Handbook for the EUROFIT tests of physical fitness*. Italia: Edigraf editorial grafica.

Da Fontoura, A. S., Schneider, P., y Meyer, F. (2004). Effect of the muscular strength detraining in prepubertal boys. *Revista Brasileira de Medicina do Esporte*, *10*, 285-288.

Delaunay, M., y Pineau, J. (1989). Un programme-la lechón-le cycle en EPS. *Revue EPS,* 217, 49-52.

DeRenne, C., Hetzler, R. K., Buxton, B. P., y Ho, K. W. (1996). Effects of training frequency on strength maintenance in pubescent baseball players. *Journal of Strength and Conditioning Research*, *10*, 8-14

Dorgo, S., King, G. A., Candelaria, N. G., Bader, J. O., Brickey, G. D., y Adams, C. E. (2009). Effects of manual resistance training on fitness in adolescents. *Journal of Strength and Conditioning Research, 23*, 2287-2294.

Faigenbaum, A. D., Milliken, L. A., LaRosa-Loud, R., Burak, B. T., Doherty, C. L., y Westcott, W. L. (2002). Comparison of 1 and 2 days per week of strength training in children. *Research Quarterly of Exercise and Sport*, *73*, 416-424.

Faigenbaum, A. D., Milliken, L. A., Moulton, L., y Westcott, W. L. (2005). Early muscular fitness adaptations in children in response to two different resistance training regimens. *Paediatric Exercise Science*, *17*, 237-248.

Hardman, K. (2008). Physical Educaton in Schools: A Global Perspective. *Kinesiology*, *40*(1), 5-28.

Hébrard, A. (1986). *L'Èducation physique et sportive. Reflexions et perspectives*. París: Coedition revue STAPS et Éditions Revue EPS.

Ingle, L., Sleap, M., y Tolfrey, K. (2006). The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *Journal of Sport Sciences*, *24*, 987-997.

Kelly, L. E., y Melograno, V. J. (2004). *Developing the physical education curriculum. An achievement-based approach*. Champaign, IL: Human Kinetics.

Le Ny, J. F. (1980). *Le conditionnement et l'apprentissage*. París: PUF.

López, R. (1992). *L'action pédagogique et formation d'adultes*. París: PUF.

Lozano, L., Viciana, J., Martínez, J. C., Cocca, A., y Jiménez, R. (2009). Influence of classroom environment and motor engagement time. *Revista Mexicana de Psicología*, *26*, 675-676.

Mayorga-Vega, D., Viciana, J., Cocca, A., y De Rueda Villén, B. (2012). Effect of a Physical Fitness Program on Physical Self-Concept and Physical Fitness Elements in Primary Schools Students, *Perceptual and Motor Skills, 115*, 3, 1-13.

Mesquita, I. (2006). Teaching well to better learn the Volleyball game. En G. Tani, J. Bento y R. Petersen (Eds.), *Sport Pedagogy* (pp. 327-343). Rio de Janeiro: Guanabara, Koogan.

Mesquita, I., Graça, A., Gomes, A. R., y Cruz, C. (2005). Examining the impact of a step game approach to teaching volleyball on student tactical decision making and skill exectution during game play. *Journal of Human Movement Studies, 48*, 469-492.

Ortega, F. B., Ruiz, J. R., Castillo, M. J., y Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity (London)*, *32*(1), 1-11.

Pesquie, P. (1988). L'Aprprentissage. *Revue EPS*, 82, 61-68.

Piéron, M. (1988). *Didáctica de las actividades físicas y deportivas*. Madrid: Gymnos.

Piéron, M. (1992). *Pédagogie des activités physiques et du sport*. París: Revue EPS.

Real Decreto 1513/2006 correspondiente a las Enseñanzas Mínimas para la Educación Primaria (2006). BOE número 293 de 6 de diciembre de 2006. Madrid.

Real Decreto 1631/2006 de Enseñanzas Mínimas para Educación Secundaria Obligatoria (2007). BOE número 5 de 7 de enero de 2007. Madrid.

Senners, P. (2001). *La lección de Educación Física*. Barcelona: INDE.

Tsolakis, C. K., Vagenas, G. K., y Dessypris, A. G. (2004). Strength adaptations and hormonal responses to resistance training and detraining in preadolescent males. *Journal of Strength and Conditioning Research*, *18*, 625-629.

Viciana, J. (1998). El aprendizaje deportivo en la edad escolar, eficacia de un tratamiento global del deporte del voleibol evaluado tradicionalmente. *Habilidad Motriz*, *11*, 36-43.

Viciana, J. (2002). *Planificar en Educación Física*. Barcelona: INDE.

Viciana, J., y Zabala, M. (2004). Un estudio descriptivo sobre cómo planifican los profesores de Educación Física. En M. A. González, A. Sánchez y J. Gómez (Ed.), *Actas del AIESEP 2002 World Congress. Preparación profesional y necesidades sociales* (pp. 732-739). A Coruña: AIESEP.

Viciana, J., Salinas, F., y Cocca, A. (2008). Effects of an Aerobic-Capacity Based Program on Cardiovascular Risk Factors in Spanish Schools. *Circulation*, *117*(11), e221.

Daniel Mayorga-Vega es Becario del programa de formación del profesorado universitario del Ministerio de Educación, Cultura y Deporte del Gobierno de España (AP2010-5905).

3. CRITERION-RELATED VALIDITY OF FIELD TESTS FOR ESTIMATING PHYSICAL FITNESS AMONG CHILDREN (PAPERS V-VIII)

Paper V

CRITERION-RELATED VALIDITY OF THE 20-M SHUTTLE RUN TEST FOR ESTIMATING CARDIORESPIRATORY FITNESS: A META-ANALYSIS

Mayorga-Vega, D., Aguilar-Soto, P., & Viciana, J.

Journal of Sports Science and Medicine

2015, *14*(3), 536-547.

©Journal of Sports Science and Medicine (2015) **14**, 536-547
http://www.jssm.org

Review article

Criterion-Related Validity of the 20-M Shuttle Run Test for Estimating Cardiorespiratory Fitness: A Meta-Analysis

Daniel Mayorga-Vega ✉, Pablo Aguilar-Soto and Jesús Viciana
Department of Physical Education and Sport, University of Granada, Spain

Abstract
The main purpose of the present meta-analysis was to examine the criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness. Relevant studies were searched from twelve electronic databases up to December 2014, as well as from several alternative modes of searching. The Hunter-Schmidt's psychometric meta-analysis approach was conducted to estimate the population criterion-related validity of the 20-m shuttle run test. From 57 studies that were included in the present meta-analysis, a total of 78 correlation values were analyzed. The overall results showed that the performance score of the 20-m shuttle run test had a moderate-to-high criterion-related validity for estimating maximum oxygen uptake (r_p = 0.66-0.84), being higher when other variables (e.g. sex, age or body mass) were used (r_p = 0.78-0.95). The present meta-analysis also showed that the criterion-related validity of Léger's protocol was statistically higher for adults (r_p = 0.94, 0.87-1.00) than for children (r_p = 0.78, 0.72-0.85). However, sex and maximum oxygen uptake level do not seem to affect the criterion-related validity values. When an individual's maximum oxygen uptake attained during a laboratory-based test is not feasible, the 20-m shuttle run test seems to be a useful alternative for estimating cardiorespiratory fitness. In adults the performance score only seems to be a strong estimator of cardiorespiratory fitness, in contrast among children the performance score should be combined with other variables. Nevertheless, as in the application of any physical fitness field test, evaluators must be aware that the performance score of the 20-m shuttle run test is simply an estimation and not a direct measure of cardiorespiratory fitness.

Key words: Maximum oxygen uptake, peak oxygen uptake, PACER, Multistage fitness test, Léger test.

Introduction

Nowadays, cardiorespiratory fitness is considered one of the most powerful markers of health, even above other traditional markers such as weight status, blood pressure or cholesterol level (Blair, 2009). Current evidence has shown how cardiorespiratory fitness status is an important quantitative predictor of cardiovascular events and all-cause mortality in healthy adults (Kodama et al., 2009). Additionally, during childhood higher cardiorespiratory fitness levels have been associated with a healthier cardiovascular profile in adulthood (Ruiz et al., 2009). Therefore, cardiorespiratory fitness testing may help to identify a target population for primary prevention both in children and adults, as well as for health promotion policies (Ruiz et al., 2009).

Different kinds of tests are commonly used to assess cardiorespiratory fitness. Cardiorespiratory fitness is typically identified as the maximal oxygen uptake (VO_2max) reached by an individual (Pescatello et al., 2014). Specifically, the VO_2max attained during a laboratory-based and graded maximal exercise test is widely considered the criterion measure (also called "gold standard") of cardiorespiratory fitness (Pescatello et al., 2014). Alternatively, due to advances in technology, today a portable gas analyzer can also be worn during a field-based graded maximal exercise test (Castagna et al., 2010; Silva et al., 2012). Due to the necessity of sophisticated and costly instrumentation, qualified technicians, and time constraints, the use of the directly measured VO_2max is limited in several settings such as in sports clubs, schools, or in large scale research studies (Pescatello et al., 2014).

Unlike the direct methods to determine VO_2max, in the above mentioned settings the performance score attained during cardiorespiratory fitness field tests could be a useful alternative. The 20-m shuttle run (20MSR) test, also called the ''Course Navette'', "PACER", or "Multistage fitness test", is probably the most widely used field test for estimating cardiorespiratory fitness (Castro-Piñero et al., 2010). The 20MSR test is simple, easy to administer and not too time-consuming, it requires minimal equipment, and a large number of individuals can be tested simultaneously. The 20MSR test consists of one-minute stages of continuous, incremental speed running. The initial speed is 8.5 km/h, and increases by 0.5 km/h per minute (Léger et al., 1984). The individual is required to run between two lines 20-m apart, while keeping pace with audio signals emitted from a pre-recorded cassette or compact disk. The test ends when the individual fails to reach the end lines concurrent with the audio signals on two consecutive occasions. Although in the original protocol the stages lasted two minutes (Léger and Lambert, 1982), later it was modified to one minute stages which were considered more motivating (Léger et al., 1984). Additionally, later different combinations of starting speed and speed increase have been proposed (e.g. Cadenas-Sánchez et al., 2014; Dong-Ho et al., 2014).

Each primary study that is published about criterion-related validity of the 20MSR test only constitutes a single piece of a constantly growing body of evidence (Cooper et al., 2009). For instance, in some studies the correlation coefficient is high (Chatterjee et al., 2006c), while in others the association is moderate or even low (Von Haaren et al., 2011). To make sense of the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses (Cooper et al., 2009; Hunter and Schmidt, 2004; Lipsey and Wilson,

Received: 04 March 2015 / Accepted: 06 May 2015 / Published (online): 01 September 2015

2001). Thus meta-analyses remain a useful tool for the evaluation of evidence (Cooper et al., 2009), forming a critical process for the development of theory in science (Hunter and Schmidt, 2004).

Previous studies have carried out meta-analyses on the validity of different field-based tests widely used in sports sciences such as the Borg's perceived exertion scale (Chen et al., 2002), the International Physical Activity Questionnaire (Kim et al., 2012), or the flexibility tests sit-and-reach (Mayorga-Vega et al., 2014a) and toe-touch (Mayorga-Vega et al., 2014b). To our knowledge there are not any meta-analyses addressing the criterion-related validity of the 20MSR test. Therefore, the purposes of the present meta-analysis were: (a) to estimate and compare the overall population mean of the criterion-related validity coefficients of the 20MSR test for estimating cardiorespiratory fitness; (b) to examine the influence of some study features (sex, age, and level of VO_2max of the participants) in criterion-related validity coefficients of the 20MSR test (between-study analyses); and (c) to compare the values of the criterion-related validity coefficients between the performance score only and the performance score combined with other variables (within-study analyses).

Methods

Search strategy

The following twelve electronic bibliographic databases were searched through December 2014: Web of Science (all databases), Scopus, SportDiscus, CINAHL, Cochrane Library Plus, ERIC, ProQuest Education Journals, Applied Social Sciences Index and Abstracts, ProQuest Social Science Journals, International Bibliography of the Social Sciences, Proquest Dissertations and Theses, and WorldCat. The searches were carried out in the search field type "Title, abstract, and keywords" or equivalent (e.g. "Topic" for the Web of Science database). Any publication format including journal papers and grey literature (i.e. master/doctoral dissertations and conference proceedings) was examined. Additionally, no language or publication date restrictions were imposed.

The search terms used were based on two concepts. Concept one included terms for the 20MSR test (navette, Léger, shuttle run, shuttle-run, shuttle test*, shuttle endurance run, multistage fitness, multi-stage fitness, beep, bleep, progressive aerobic cardiovascular endurance, PACER, 20 m test*, 20-m test*, 20 m run, 20-m run, bip test*) and concept two included terms related to validity (valid*, related, relationship, correlation, regression, comparison, association, estimat*, determinat*, predict*, equation*, VO_2*, oxygen uptake, oxygen intake, consumption of oxygen, oxygen consumption, aerobic, cardiovascular, cardiorespiratory, fitness, gold standard, criterion measur*). The truncated root of certain terms was followed by an asterisk to include multiple variants. Additionally, the keywords that consisted of more than one word were enclosed in quotes. Finally, the terms of the same concept were combined together with the Boolean operator "OR" and then the two concepts were combined using the Boolean operator ''AND'' (Cooper et al., 2009).

Based on the results of the Boolean-based search (as well as all the related studies by Léger), other modes of searching were carried out. The reference lists of all studies (as well as some related studies reviews) were manually searched (also called "snowballing"). Additionally, the reference citations (in the Web of Science and Scopus databases) and the researcher publications of the first authors (in the Web of Science, Scopus and SportDiscus databases) were also examined. Subsequently, the authors for correspondence (if they were not defined, the first author was used) were contacted by email. Finally, the researcher's personal lists (in ResearchGate, Google Scholar, and personal websites) of the first authors were screened. Any time a new record was found, all of these modes of searching were repeated until any new study appeared.

Selection criteria

The selection criteria to identify studies that examined the criterion-related validity of the 20MSR test were: (1) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (2) studies with the original protocols of the 20MSR test (Léger and Lambert, 1982; Léger et al., 1984) or some modifications of them in starting speed, speed increase and/or duration of stages (also called "levels" or "paliers"); (3) studies in which for the criterion measure the VO_2max was measured in a standardized and laboratory-based incremental test to exhaustion; and (4) studies which associated the performance scores of the field test (or the performance score with other variables) with the measured VO_2max results using a Pearson's r zero-order correlation coefficient or simple linear regression (R^2) (or a multiple linear regression in case of multiple predictors).

Coding studies

For the present meta-analysis, from each selected study the following data were coded: Identification number, type of publication (1 = journal paper, 2 = grey literature –master dissertation, doctoral dissertation, or conference proceeding-), sample size (n), sex of participants (1 = men, 2 = women, 3 = men and women, 4 = no information), age of participants (1 = children, < 18 years; 2 = adults, ≥ 18 years; 3 = children and adults; 4 = no information), 20MSR test protocol (1 = Léger's protocol, 2 = Eurofit protocol, 3 = QUB's protocol; 4 = others), criterion measure protocol (1 = treadmill run test; 2 = cycle ergometer test; 3 = others), measurement unit of the 20MSR test (1 = completed stage, accuracy of one; 2 = completed stage, accuracy of half; 3 = total laps; 4 = speed expressed in km/h; 5 = time expressed in seconds; 6 = distance expressed in metres; 7 = others), measurement unit of the criterion test (1 = VO_2max expressed in ml/kg/min; 2 = VO_2max expressed in l/min; 3 = maximal aerobic speed expressed in km/h; 4 = other), mean value of the measurement criterion, reliability of the 20MSR test (intraclass correlation coefficient), reliability of the measurement criterion (intraclass correlation coefficient), statistical test used for the criterion-related validity (1 =

Pearson's *r* correlation coefficient, 2 = R^2 simple or multiple linear regression), and criterion-related validity value (separately for performance score only and multiple predictors). In addition, any observations were also registered when some special question was found.

Although various protocols for evaluating study quality have been described, there is no widespread agreement on the validity of this kind of evaluation approach (e.g. see Cooper et al., 2009). Thus, rejecting certain studies and accepting others for inclusion in a meta-analysis on the basis of a quality score remains a controversial procedure (Cooper et al., 2009; Flather et al., 1997). Therefore, according to Flather et al. (1997), in the present meta-analysis the approach followed has been to ensure that the design has not been flawed (e.g. the VO_2max was measured in a standardized and laboratory-based incremental test to exhaustion), and that there has been a complete reporting of relevant outcomes. For a study to be included in this meta-analysis, sample size, protocol of the 20MSR test, unit and protocol of the criterion measure test, statistical test, and value of the criterion-related validity were considered to be critical. In the event that the authors failed to identify any study feature, they were contacted to retrieve it. If the study feature was not retrieved, the data was omitted. If the data missed any critical value, the study was not included in the meta-analysis.

Studies selection criteria were examined by two independent researchers. However, because the identification of the features of a study is usually explicitly stated in primary papers, data were coded by only one researcher (except for the criterion-related validity values that were coded by two independent researchers). When doubt or disagreement occurred, a consensus was achieved through discussion.

Data analyses

According to Hunter and Schmidt (2004), in the present meta-analysis Pearson's zero-order correlation coefficient (*r*) was considered the unit of measurement as an indication of the criterion-related validity of the 20MSR test. When the validity values were reported as R^2, therefore, it was transformed. Additionally, to avoid dependency issues in the meta-analysis, an exhaustive examination of the selected studies was carried out. All the examined studies used the relative VO_2max (i.e. expressed in ml/kg/min) as the measurement criterion. Although some studies also reported criterion-related validity results using additional markers such as the absolute VO_2max expressed in l/min (Aziz et al., 2007; McIver et al., 2004), relative VO_2max using lean body mass (Varness et al., 2009), or the maximal aerobic speed (Kuisis, 2007), these validity coefficients were not selected. Since some studies used multiple performance scores of the 20MSR test for examining the criterion-related validity (LaMontagna, 1991; Matsuzaka et al., 2004; Metsios et al., 2006; Ramsbottom et al., 1988; Suminski et al., 2004; Varness et al., 2009), the average value was used. Nevertheless, when authors reported the results of criterion-related validity from the combination of different multiple predictors (Barnett et al., 1993; Mahar et al., 2006; 2011; Hamlin et al., 2014), only the best model (i.e. higher coefficient value) was used in the present meta-analysis.

If a single study reported more than one *r* value within the same 20MSR test protocol, but from different subsamples, we assumed each *r* value from different subsamples to be independent and included them in a single meta-analysis (Lipsey and Wilson, 2001). When, in the same study, data for men and women were expressed both separately and together, only the separate data were selected (e.g. Hamlin et al., 2014; Silva et al., 2012; Von Haaren et al., 2011). However, when in the same study, data for the whole and subsamples with respect to sex and age categories were expressed, only the whole sample was coded (e.g. Mahar et al., 2006). Similarly, when in the same study, data were expressed for different days from the same sample (i.e. LaMontana, 1991; McIver et al., 2004), the average value of the coefficients was coded.

Publication bias: In addition to the search strategy followed and selection criteria to avoid availability bias, another examination of the selected studies was carried out to avoid a potential duplication of information retrieved. Similarities between studies of the same authors, with the same correlation coefficients and/or the same sample size were examined. If some selected studies had full or partial duplicated information, these particular correlation values were not analyzed in the meta-analyses. Furthermore, before computing correlations, several exploratory analyses were also conducted for identifying and assessing the impact of any potential publication bias. Firstly, according to Light and Pillemer (1984), the scatter plots of correlation coefficients against sample size for each 20MSR test protocol were analyzed. Secondly, with the objective of quantifying the outcomes of the scatter plots, based on Begg and Mazumdar (1994), a Spearman's rank order correlation between *r* values and sample size was calculated. Finally, for assessing the impact of any potential publication bias, a file drawer analysis based on effect size was performed to estimate the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean effect size (r_p) down to the small mean *r* value (Orwin, 1983). According to Cohen's (1992) guidelines, the correlation coefficient was interpreted as small when $r < 0.30$.

Computation of correlations: The Hunter-Schmidt's psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of the 20MSR test (Hunter and Schmidt, 2004). This approach estimates the population correlation correcting the observed correlations due to various artefacts such as sampling error and measurement error. The "bare-bone" mean *r* (r_c), corrected for only sampling error was first calculated by weighting each *r* with the respective sample size when aggregating them into r_c. Then, we calculated the corrected mean *r* at the population level (r_p) that was unaffected by both sampling error and measurement error. Since the reliability coefficients (intraclass correlation coefficients) of the 20MSR were unavailable in most of the included primary studies, the measurement error was corrected using artifact distributions instead of individually. On the other hand, the measurement error of the criterion test could not be corrected because the reliability was not available. Finally, the 95% confidence intervals of r_p (95% CI) were calculated.

Moderator analysis: In the present meta-analysis, due to the low number of r values found, partial hierarchical analyses of moderator variables were carried out. According to Hunter and Schmidt (2004), to determine the presence of moderator effects which may affect overall criterion-related validity of the 20MSR test (r_p), three different criteria were simultaneously examined: (a) the 95% credibility interval (95% CV) is relatively large or includes the value zero; (b) the percentage of variance accounted for by statistical artefacts is less than 75% of the observed variance in r_p; and (c) the Q homogeneity statistic is statistically significant (p < 0.05). If at least one of the three criteria were met, we concluded that the results could be affected by moderator effects. In the presence of moderator effects, criterion-related validity values of the 20MSR test were analyzed separately by: (a) sex of participants (i.e. men and women); (b) age of participants (i.e. children and adults); and (c) level of VO_2max (i.e. low average level, < P_{50}, and high average level, ≥ P_{50}) (between-study analyses). Additionally, the criterion-related validity values of the 20MSR test for the performance score only were compared with the criterion-related validity with multiple predictors (within-study analysis).

The meta-analyses were performed using the software Hunter and Schmidt Meta-Analysis Programs version 1.1 for Windows (Iowa, 2005). All the others statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20).

Results

Study description

Figure 1 shows a flowchart of the study selection process. Of the 7,777 bibliographic databases search results, 238 potentially relevant publications were identified and retrieved for a more detailed evaluation. Afterward, based on the 52 studies of the Boolean-based search that met the selection criteria (plus 16 studies reviews that were also used for the reference lists mode), other modes of searching were carried out. Through the other modes of searching eight additional studies met the selection criteria. However, due to duplication, of the overall 60 studies that met the inclusion criteria, 57 studies were included in the present meta-analysis. Finally, from the 57 studies that were included in the present meta-analysis, a total of 78 r values across three 20MSR test protocols were retrieved, being 65 r correlation coefficients for the criterion-related validity using the performance score only and 13 r correlation coefficients for multiple predictors (i.e. the performance score and other variables: age, sex, biological maturation, body mass, body mass index, body fat and/or skinfolds).

In the present meta-analysis 54 studies with performance score only (Aandstad et al., 2011; Armstrong et al., 1988; Aslan et al., 2012; Aziz et al., 2005b; 2007; Bandyopadhyay, 2011; 2013; Barnett et al., 1993; Chatterjee et al., 2005; 2006a; 2006c; 2008a; 2008b; 2008c, 2009; 2010a; 2010b; 2010c; 2011; 2013; De Souza et al., 2010; Dickau, 2011; Dong-Ho et al., 2014; Flouris et al., 2004; 2006; Gadoury and Léger, 1986; Green et al., 2013; Hamlin et al., 2014; Kuisis, 2007; LaMontagna, 1991; Liu et al., 1992; Mahar et al., 2002; 2006; 2011; 2013; Mahoney, 1992; Matsuzaka et al., 2004; McIver et al., 2004; McVeigh et al., 1995; Metsios et al., 2006; Mombiedro et al., 1992; O'Gorman et al., 2000; Paliczka et al., 1987; Paradisis et al., 2014; Pitetti et al., 2002; Poortmans et al., 1986; Ramsbottom et al., 1988; Stickland et al., 2003; Suminski et al., 2004; Thomas et al., 2006; Van Mechelen et al., 1986; Van Praagh et al., 1988; Varness et al., 2009; Von Haaren et al., 2011) and 11 studies with multiple predictors were included (Barnett et al., 1993; Chia et al., 2005; Dong-Ho et al., 2014; Hamlin et al., 2014; Mahar et al., 2002; 2006; 2010; 2011; McVeigh et al., 1995; Matsuzaka et al., 2004; Tsiaras et al., 2010).

Some studies retrieved for a more detailed evaluation were not included because they were carried out with non-healthy participants (e.g. individuals with Down's syndrome, cerebral palsy or in wheelchairs) (e.g. Agiovlasitis et al., 2011; Goosey-Tolfrey and Tolfrey, 2008; Kloyiam et al., 2011), used mayor modifications of the 20MSR test (e.g. the Square shuttle run test or Yo-Yo intermittent recovery test) (e.g. Castagna et al., 2008). Other studies that were retrieved for a more detailed evaluation were not selected because only cross-validity was examined (e.g. Batista et al., 2013). Not one of three potential studies were selected because they did not define the protocol used (i.e. lacked critical information) and the authors did not reply when asked for it (Cunningham et al., 1994; Hemmings et al., 2003; Lightburne, 2008). Then, the full-text of some potential studies was not available (Barnejee et al., 2005; Chartterjee et al., 2007).

Finally, some potential studies were not selected because they did not use the measured VO_2max during a standardized and laboratory-based incremental test to exhaustion as a criterion measure. For instance, some research studies assessed the VO_2max during the field test (e.g. Castagna et al., 2010; Silva et al., 2012). Nevertheless, previous studies have found that the measured VO_2max during the 20MSR test is significantly different compared with that measured during a laboratory-based test (Aziz et al., 2005a; Flouris et al., 2010). In other potential studies (e.g. Léger and Lambert, 1982; Léger et al., 1988) the VO_2max was assessed by retroextrapolating the O_2 recovery curve at time zero of recovery. Retroextrapolation is a method to estimate VO_2max (i.e. indirect measure) and, therefore, it cannot be considered as a criterion measure to determine it (i.e. direct measure) such as assessing the VO_2max during a standardized incremental test to exhaustion (Aslan et al., 2012; Mahar et al., 2011).

Publication bias

Firstly, several exploratory analyses were followed to avoid full or partial duplicated information availability bias. Although three research studies met the selection criteria (Chatterjee et al., 2006b; Hamlin et al., 2013; Paradisis et al., 2013), the correlation coefficient value was not analyzed in the present meta-analyses. Paradisis

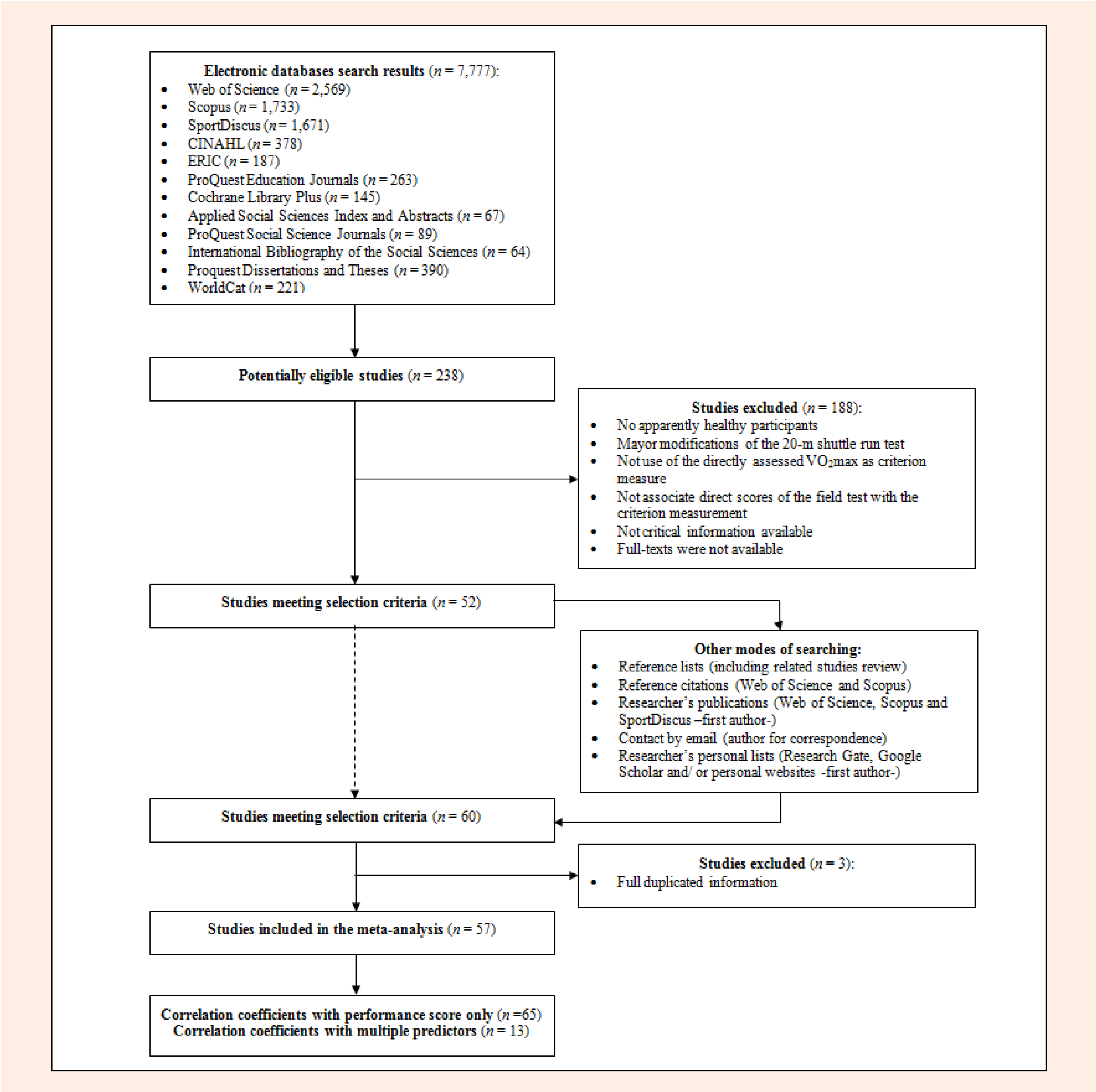


Figure 1. Flowchart of the study selection process.

et al. (2013) and Hamlin et al. (2013) conference papers were not included because they were published later in a journal (Hamlin et al., 2014; Paradisis et al., 2014). Actually, Paradisis and colleagues (2013) in their conference paper published the results of a pilot study using a subsample. Regarding the study by Chatterjee et al. (2006b), the same study had been also published in another journal (Chatterjee et al., 2008c).

Exploratory analyses were conducted to identify the presence of publication bias. Because the sum of the r values for some protocols was very small, the following analyses were calculated only for Léger's (for both performance score only and multiple predictors) and Eurofit protocols (for performance score only). Figure 2 shows the scatter plots of sample size against criterion-related validity coefficients for estimating VO_2max for Léger's (for performance score only and multiple predictors) and Eurofit protocols (for performance score only). According to this graphical method, the figures suggest that for the performance score only there was not publication bias for both protocols. Nevertheless, we have to be aware that for Léger's protocol seems to be a slightly major density in the right-hand corner. For Léger's protocol with multiple predictors, however, the scatter plot suggests the presence of publication bias, because of the absence of r values in the lower left hand corner.

The results of Spearman's rank order correlation between r values and sample size did not show any statistically significant correlation (Léger's protocol, performance score only: r = 0.08, p = 0.601; multiple predictors: r = - 0.26, p = 0.450; Eurofit protocol, r = - 0.06, p = 0.873). Due to the small number of rs found for the Léger's protocol (with multiple predictors) and the Eurofit protocol, the results of both the scatter plot and the Spearman correlation must be interpreted with caution (Begg and Mazumdar, 1994; Cooper et al., 2009). Additionally, empirical evaluations of the funnel plots suggest that their interpretation can be limited (Terrin et al., 2005).

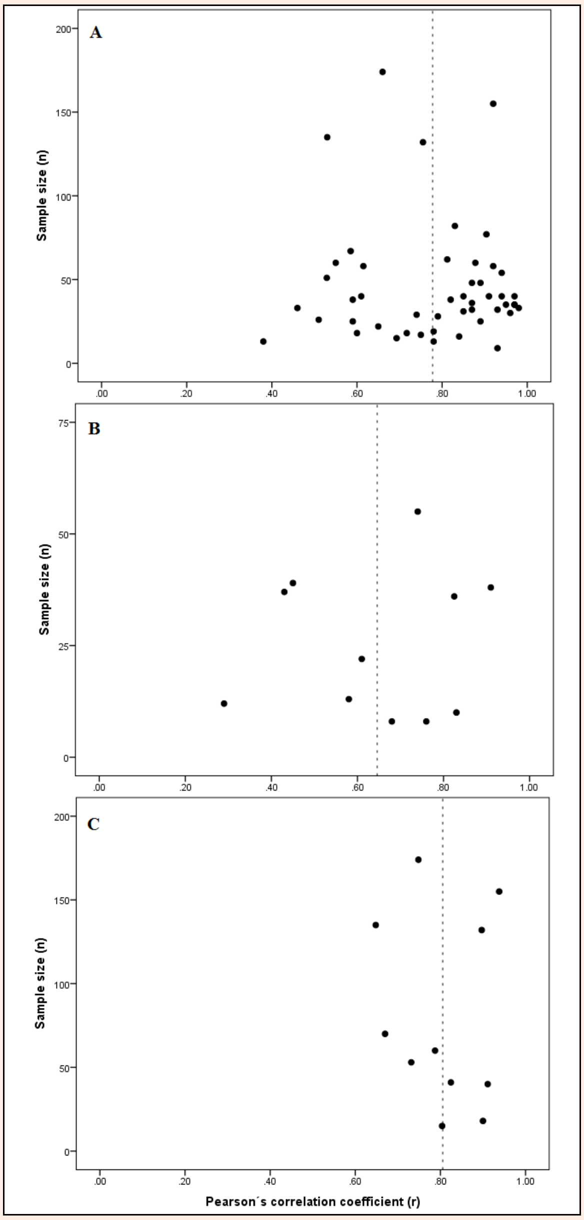


Figure 2. Scatter plot of sample size against criterion-related validity coefficients (*r*) for estimating maximal oxygen uptake: (a) Léger's protocol with performance only score; (b) Léger's protocol with multiple predictors; and (c) Eurofit protocol with performance score only. Dashed line represents mean values of validity coefficients.

Finally, file drawer analyses based on effect size were carried out for assessing the impact of any potential publication bias. The results of the file drawer analyses are based on effect size for estimating the number of unlocated studies averaging null results (r = 0) that would have to exist to bring the mean r_p down to 0.29. These results are shown in the following lines (in parenthesis the unlocated/located percentage): for performance score only, Léger's protocol 80 (167%), Eurofit protocol 14 (127%), QUB's protocol 7 (140%), and Do-Hong's protocol 2 (200%); for multiple predictors, Léger's protocol 20 (182%), Eurofit protocol 2 (200%), and Do-Hong's protocol 2 (200%). Although we are aware that there is not a large number of "lost" studies for some protocols, the results for the percentage of unlocated/located studies showed an unlikely number of "lost" studies (127-200%).

Criterion-related validity

Table 1 reports the number of r values studied (K), the total sample size accumulated (N), the overall weighted mean of r corrected for sampling error only (r_c), the overall weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI for overall criterion-related validity correlation coefficients (r_p) for estimating VO_2max across each 20MSR protocol. Additionally, to detect the presence of moderator effects which may affect overall criterion-related validity of the 20MSR test, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the Q homogeneity statistic were calculated.

The overall results showed that the 20MSR test had a moderate-to-high mean correlation coefficient of criterion-related validity for estimating VO_2max in which all 95% CI did not include the value zero. The results of the present meta-analysis also showed that the criterion-related validity of Léger's protocol was statistically higher than the QUB's (Queen's University Belfast) protocol. For Léger's and Eurofit protocols the percentage of variance accounted for by statistical artefacts was less than 75%, and the 95% CV was relatively large, as well as for Léger's protocol the Q homogeneity statistic was also statistically significant (p < 0.05). Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study. However, since none of the three criteria were met in the QUB's protocol, moderator analyses were not conducted for that particular protocol.

Regarding the multiple predictors, the overall results showed that when the performance score of the 20MSR test was combined with other variables the mean correlation coefficients of criterion-related validity for estimating VO_2max were moderate-to-very-high. Additionally, when the 95% CI could be calculated (i.e. for Léger's protocol), the value zero was not included. Although two of the three criteria were met in Léger's protocol, due to the small n for the most of its subcategories (e.g. only one correlation coefficient for adults subcategory and two for men subcategory), the between-study moderator analyses were not conducted in that case. However, due to the fact that for Léger's protocol eight studies reported correlation coefficients of criterion-related validity for both performance score only and combined with other variables, the within-study analysis was conducted as it was hypothesized in the present study (see moderator analyses).

Moderator analyses

Table 2 shows the results of between-study moderator analyses to examine the effects of sex (i.e. men and women), the age of participants (i.e. children and adults), and

Table 1. Results of meta-analyses for overall criterion-related validity correlation coefficients across the 20-m shuttle run test protocols.

Protocols	*K*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% variance[c]	*Q* statistic
Performance score only								
Léger[d]	48	2,222	.77	.84	.80-.89	.54-1.00	15.15	313.42*
Eurofit[e]	11	278	.66	.73	.60-.86	.43-1.00	42.50	18.10
QUB[f]	5	401	.68	.71	.64-.77	.64-.77	79.86	1.34
Dong-Ho[g]	1	127	.62	.66	-	-	-	-
Performance score with other variables								
Léger[d]	11	893	.80	.87	.81-.94	.67-1.00	14.94	73.04*
Eurofit[e]	1	55	.85	.95	-	-	-	-
Dong-Ho[g]	1	127	.73	.78	-	-	-	-

Note. *K*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error of the 20-m shuttle run test; [a]95% confidence interval; [b]95% credibility interval; [c]Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of the 20-m shuttle run test; [d]Léger's protocol starts at 8.5 km/h and increases 0.5 km/h each minute (Léger et al., 1984, 1988); [e]Eurofit protocol starts at 8.0 km/h and increases 0.5 km/h each minute, but the second stage increases by 1.0 km/h (Council of Europe Committee for the Development of Sport, 1988); [f]QUB's protocol starts at 8.0 km/h and increases 0.5 km/h each minute (Riddoch, 1990); [d]Dong-Ho's protocol starts at 7.5 km/h and increases 0.5 km/h each minute (Dong-Ho et al., 2014). * $p < 0.05$

Table 2. Results of moderator analyses for criterion-related validity correlation coefficients across the 20-m shuttle run test protocols potentially affected by moderator effects†

Moderator	Effect	*K*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% variance[c]	*Q* statistic
Between-study analyses									
Sex of participants									
Léger[d]	**Men**	24	782	.80	.88	.81-.94	0.59-1.00	18.35	124.53*
	Women	13	475	.74	.81	.70-.92	0.50-1.00	21.56	55.14*
Eurofit[e]	**Men**	6	125	.61	.68	.49-.87	0.41-0.95	57.29	5.44
	Women	4	98	.67	.75	.53-.97	0.38-1.00	32.74	10.00*
Age of participants									
Léger[d]	**Children**	28	1,335	.72	.78	.72-0.85	0.50-1.00	22.30	113.74*
	Adults	20	887	.86	.94	.87-1.00	0.72-1.00	12.72	160.02*
Eurofit[e]	**Children**	7	143	.61	.68	.52-.84	0.50-0.86	76.18	2.66
	Adults	4	135	.71	.79	.56-1.00	0.43-1.00	24.49	15.00*
Level of VO_2max									
Léger[d]	**Low**	22	1,181	.75	.82	.74-.89	0.49-1.00	13.27	167.60*
	High	23	895	.80	.88	.80-.95	0.62-1.00	18.84	115.48*
Eurofit[e]	**Low**	5	108	.69	.77	.59-.94	0.41-1.00	36.29	10.68*
	High	6	170	.64	.71	.52-.91	0.45-0.98	48.32	7.81
Within-study analysis									
Number of predictors[f]									
Léger[d]	**One**	8	742	.70	.77	.68-.86	0.51-1.00	15.56	50.61*
	Few	8	742	.80	.88	.80-.95	0.67-1.00	12.60	64.68*

Note. *K*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error of the 20-m shuttle run test; VO_2max, maximal oxygen uptake; [a]95% confidence interval; [b]95% credibility interval; [c]Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of the 20-m shuttle run test; [d]Léger's protocol starts at 8.5 km/h and increases 0.5 km/h each minute (Léger *et al*., 1984, 1988); [e]Eurofit protocol starts at 8.0 km/h and increases 0.5 km/h each minute, but the second stage increases by 1.0 km/h (Council of Europe Committee for the Development of Sport, 1988); [f] Performance only score ("one") or performance score plus other ("few"). †Because some studies mixed categories or some values were missing, the overall *n* for some categories is lower for some 20-m shuttle run tests. * $p < 0.05$

the level of VO_2max (i.e. low average level, < P_{50} and high average level, ≥ P_{50}) on overall criterion-related validity correlation coefficients for estimating VO_2max for each 20MSR protocol potentially affected by moderator effects (i.e. Léger's and Eurofit protocols). Additionally, for Léger's protocol the correlation coefficients of criterion-related validity which were reported for both the performance score only and multiple predictors were compared (i.e. within-study analysis).

Sex of participants: The results showed that the analyzed 20MSR protocols had a moderate-to-high mean correlation coefficient of criterion-related validity for estimating VO_2max for both men and women in which all 95% CI did not include zero. Moreover, all the 95% CI of mean correlation coefficients overlapped. On the other hand, we must point out that, according to moderator analysis criteria, at least two of the three criteria were met in the 20MSR test, indicating that the criterion-related validity of these protocols separately for sex was still heterogeneous. Finally, because some studies grouped men and women together or the sex values were missing, in Table 2 overall *n* for sex of participants is lower.

Age of participants: Results showed that Léger's protocol had a moderate mean correlation coefficient of criterion-related validity for estimating VO_2max for children and moderate-to-high for adults. Additionally, the results of the present meta-analysis showed that the criterion-related validity of Léger's protocol was statistically higher for adults than for children. On the contrary, the results showed that the Eurofit protocol had a moderate mean correlation coefficient of criterion-related validity for estimating VO_2max for both children and adults, as well as the 95% CI of mean correlation coefficients overlapped. In addition, the 95% CIs included the value zero. Finally, according to moderator analysis criteria, at least one of the three criteria was met in all categories, indicating that the criterion-related validity of these protocols separately for age was still heterogeneous.

Level of VO_2max: The results showed that the analyzed 20MSR protocols had a moderate-to-high mean correlation coefficient of criterion-related validity for both participants with low and high level of VO_2max in which all 95% CI did not include the value zero. Furthermore, all the 95% CI of mean correlation coefficients overlapped. Regarding the moderator analysis criteria, at least two of the three criteria were met in all categories, indicating that the criterion-related validity of these 20MSR protocols separately for level of VO_2max were still heterogeneous. Finally, because for Léger's protocol some studies failed to identify the level of VO_2max, in Table 2 overall *n* for level of VO_2max is lower.

Number of predictors: The results showed that for Léger's protocol the performance score only had a moderate mean correlation coefficient of criterion-related validity for estimating VO_2max meanwhile, when the performance score was added to other variables the criterion-related validity was moderate-to-high. However, the 95% CI of mean correlation coefficients overlapped. Additionally, the three criteria were met in both categories, indicating that the criterion-related validity was heterogeneous.

In summary, the overall results showed that the 20MSR test had a statistically significant and moderate-to-high criterion-related validity for estimating VO_2max. Regarding the moderator analyses, the criterion-related validity of the 20MSR test was statistically higher for adults than for children; however, sex and VO_2max levels did not seem to affect the criterion-related validity of the 20MSR test.

Discussion

The first purpose of the present meta-analysis was to estimate and compare the overall population mean of the criterion-related validity coefficients of the 20MSR test for estimating cardiorespiratory fitness. The choice of a cardiorespiratory fitness test must be based on its functionality and validity. Although the VO_2max measured during a laboratory-based and graded maximal exercise test has the advantage of being the criterion measure to assess cardiorespiratory fitness, due to several practical reasons they have the disadvantage of having a limited use in several settings (Pescatello et al., 2014). In settings such as sports clubs, schools or large scale research studies, as the 20MSR test has the advantage of allowing for an evaluation in a short amount of time with minimal skill and instrumentation, potentially it could be a useful alternative to estimate cardiorespiratory fitness. In this context, the overall results of the present meta-analysis show that the 20MSR test has a moderate-to-high mean correlation coefficient of criterion-related validity for estimating VO_2max.

Since the original 20MSR test of one-min stages (Léger et al., 1984), various modifications of the start and subsequent speed increases have been proposed (e.g. Council of Europe Committee for the Development of Sport, 1988; Dong-Ho et al., 2014; Riddoch, 1990). However, according to the results of the present meta-analysis (and despite the fact that we are aware that the 95% CI of mean correlation coefficients with the Eurofit protocol overlapped), Léger's protocol showed a greater average criterion-related validity coefficient. Therefore, if our purpose is to assess cardiorespiratory fitness, it seems that the use of the Eurofit and QUB's protocols is not justified. However, the fact that in the present meta-analysis the overall criterion-related validity of Léger's and Eurofit protocols was heterogeneous, as well as the number of *n* for the QUB's protocol was low, must be highlighted. Additionally, any primary study comparing the criterion-related validity of various protocols of the 20MSR test was found. Therefore, we should be cautious with the overall results of the present meta-analysis.

The second purpose of this meta-analysis was to examine the influence of some potential moderator factors (sex, age, and level of VO_2max of the participants) in criterion-related validity coefficients of the 20MSR test. One of the main findings of the present meta-analysis showed that the criterion-related validity of Léger's protocol was significantly higher for adults compared with children. Similarly, for the Eurofit protocol the average correlation coefficient for adults was considerably higher than for children. Although we have to be aware that for that protocol the 95% CI overlapped, the large CI probably because of the low number of correlations found must also be considered. Additionally, while 51% of the total simple correlation coefficients with children were equal or below 0.70 (i.e. less than 50% of variance explained), only 20% was found for adults. Therefore, the results of the present meta-analyses show that the criterion-related validity of the 20MSR test is statistically higher for adults versus children.

In line with the findings of the present meta-analysis, Matsuzaka et al. (2004) found out that, when participants were examined under the same experimental conditions (e.g. field and laboratory tests protocols, equipment, and testers), the criterion-related validity of the 20MSR test was considerably higher for adults (r = 0.92) than for children (r = 0.75) and adolescents (r = 0.76). Similarly, Léger et al. (1988) suggested that, since the chronological age of children, and not adults, was a significant predictor of VO_2max, the lower validity of the 20MSR test in children as compared to adults might be the result of larger interindividual variations. In addition to chronological age, in the present meta-analysis, it has been found that among children other variables such as sex, biological maturation, body mass, body mass index, body fat and/or skinfolds were significant predictors of the VO_2max. Furthermore, since children might be less willing to endure discomfort of strenuous effort, have less motivation, and/or a limited attention span for monotonous tasks, the 20MSR test performance could be affected and, therefore, its criterion-related validity.

Another potential reason for these results could be that the starting speed of the 20MSR test is too high for children. Current evidence suggests that to elicit valid VO_2max values, continuous incremental tests should last at least five minutes (Midgley et al., 2008). However, Castro-Piñero et al. (2011) in a population-based study carried out using Léger's protocol (i.e. starting speed 8.5 km/h) found that most 6-to-17-year-old children did not complete five stages (i.e. five minutes). Previous studies have proposed modifications of the 20MSR test for children with a drastically reduced starting speed (e.g. 4 km/h, Quinart et al., 2014; 6.5 km/h, Cadenas-Sánchez et al., 2014). Unfortunately, these authors did not either examine the criterion-related validity of the test (Cadenas-Sánchez et al., 2014) or did not compare it with "traditional" protocols such as the Léger protocol (Quinart et al., 2014). As regards the moderator analyses for sex and VO_2max levels, according to the results of the present meta-analysis they seem not to affect the criterion-related validity. Therefore, since the criterion-related validity for both men-women and low-high level of VO_2max subgroups was similar, the 20MSR test can be used interchangeably for any subcategory. Regarding the VO_2max categories, however, due to the fact that in the present meta-analysis the *n* was classified based on the average score, we have to be aware that several participants with low VO_2max values could be classified as high values and vice versa. This fact could affect the results of the present meta-analysis.

Finally, the third purpose of the present meta-analysis was to compare the values of the criterion-related validity coefficients between the performance only score and the performance score combined with other variables. When multiple predictors were used, the average correlation coefficient was considerably higher than for the performance score only (on average $r\Delta = 0.11$). Although we have to be aware that for that protocol the 95% CI overlapped, the large CI, probably because of the low number of correlations found must also be considered. It must be also pointed out that seven of the eight studies analyzed were carried out with children. Thus, children seem to benefit considerably from other variables to estimate cardiorespiratory fitness. In summary, although the criterion-related validity of the 20MSR test is statistically lower for children than for adults, when the performance score is combined with other variables such as age, sex, body mass or body mass index, the criterion-related validity value of the 20MSR test is considerably high.

Strengths and limitations

The meta-analysis is a useful tool to assess scientific evidence, but an understanding of its strengths and limitations is needed for the most appropriate use of this method. An extensive revision of the general strengths and limitations of meta-analyses (e.g. Cooper et al., 2009), as well as specifically in the meta-analysis of the criterion-related validity of physical fitness field tests has been published elsewhere (Mayorga-Vega et al., 2014a).

Briefly, regarding the strengths of the present meta-analysis, we followed several measures to avoid (or at least to reduce) publication bias. Firstly, to avoid availability bias, we conducted a wide literature search through several databases without limiting any kind of manuscript, language or publication date. Due to the limitations of databases to find the "fugitive" literature, several complementary searches were also carried out. Secondly, in the present meta-analysis all the studies published by the same authors were thoroughly cross-referenced with each other in order to avoid duplication. Lastly, several exploratory analyses were also conducted to identify and assess the impact of any potential publication bias.

Another strength of the present meta-analysis is related to the statistical approach used. In the present study, the Hunter-Schmidt's psychometric meta-analysis approach (2004) was conducted in order to obtain the population estimates of criterion-related validity of the 20MSR test. Since this method estimates the population correlation by correcting the observed correlations due to various artefacts such as sampling error and measurement error, it has been considered one of the best meta-analyses approaches.

Regarding the limitations, there were some that should be considered when examining the results of the present meta-analysis. The main limitations were related to the small number of criterion-related validity coefficients found. Estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis. Because a partial hierarchical breakdown (instead of full) had to be used, misleading results due to confounding and interaction effects might be produced (Hunter and Schmidt, 2004). Therefore, the results of the present study should be considered with caution; firmer conclusions should await the accumulation of a larger number of studies (Hunter and Schmidt, 2004).

Another limitation of the present meta-analysis is related to the criterion measure used in the studies. Although in the present meta-analysis only primary studies that used as the criterion measure, the VO_2max during a standardized and laboratory-based incremental test to exhaustion were selected (see results section), in these studies different equipment (various brand and characteristics), ergometers (i.e. treadmill and cycle ergometer) and protocols (e.g. in warm-up, initial load, increasing load, gas collection time, or number of gas collections) were used. Furthermore, in the studies there is not a wide agreement about the criteria to determine VO_2max. For example, researchers used a plateau in VO_2, the respiratory exchange ratio, or the age-adjusted estimates of the maximal heart rate, alone or in combination; then, the quantitative cut-off values criteria are also diverse.

The fact that in the present meta-analysis peak oxygen uptake (VO_2peak) has been used interchangeably with VO_2max must be highlighted. Although we are aware that the VO_2peak simply refers to the highest value of oxygen uptake (VO_2) attained on a particular exercise test, due to the fact that the tests in the primary studies of the present meta-analysis were maximal we can be reasonably sure that values were the highest value of VO_2 that is deemed attainable by individuals, i.e. the VO_2max (Rowland, 1993). Therefore, it seems that the criterion measure of cardiorespiratory fitness should be reexamined and readjusted (Howley et al., 1995).

Finally, coding some study features was problematic due to different reasons. For instance, because in the present meta-analysis the level of VO_2max was classified based on the average scores, we are aware that several individuals with low VO_2max could be classified as high VO_2max and vice versa. Additionally, although participant characteristics such as physical activity levels or sport participation were potentially moderating features, coding for them was not possible because most studies did not identify them.

Conclusion

Overall the 20MSR test has a moderate-to-high mean correlation coefficient of criterion-related validity for estimating VO_2max. Regarding the potential moderators examined, the present meta-analysis shows that the criterion-related validity of the 20MSR test is higher for adults than for children. Nevertheless, when the performance score among children is combined with other variables, the criterion-related validity to estimate the VO_2max is considerably high. As regards the sex and level of VO_2max of participants, they seem not to affect the relationship between the 20MSR test score and the measured VO_2max.

When an individuals' VO_2max attained during a laboratory maximal exercise test is not feasible such as in sports clubs, schools or large scale research studies, scientists and practitioners could use the 20MSR test as a useful alternative to estimate cardiorespiratory fitness. Among adults the performance only score seems to be a strong estimator of cardiorespiratory fitness, in contrast among children the performance score should be combined with other variables. Nevertheless, as in the application of any physical fitness field test, testers must be aware that the performance score of the 20MSR test is simply estimation and not a direct measure of cardiorespiratory fitness.

Due to the relatively low number of *r* values found and that criterion-related validity of the 20MSR test within most categories is heterogeneous, we should be cautious with the results of the present meta-analysis. Therefore, when a greater number of studies are accumulated, a large sized meta-analysis with a full hierarchical analysis approach should be carried out. For this purpose future research studies should further examine the criterion-related validity of the 20MSR test, especially in modifications of the test with a lower starting speed, among populations such as children, and go deeper into other related aspects such as the potential moderator effects of the level of VO_2max.

Acknowledgements

We gratefully acknowledge all the authors of the original research studies for their contribution, without whom the present meta-analysis could not be done. We especially acknowledge Professor Luc Léger for his gracious help in the identification of several synonyms for the test. Furthermore, we thank the head of the library, Mrs. Ana M. Peregrín González, for technical assistance and her great help to retrieve manuscripts that were not readily available. Additionally, we would like to express our gratitude to the Associate Editor-in-Chief Roger Ramsbottom for all the hard work and detailed care that have undoubtedly improved the present manuscript. We also thank Aliisa Hatten-Viciana for the English revision. Daniel Mayorga-Vega is supported by the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

References

Aandstad, A., Holme, I., Berntsen, S. and Anderssen, S.A. (2011) Validity and reliability of the 20 meter shuttle run test in military personnel. *Military Medicine* **176,** 513-518.

Agiovlasitis, S., Pitetti, K.H., Guerra, M. and Fernhall, B. (2011) Prediction of VO_2peak from the 20-m shuttle-run test in youth with Down syndrome. *Adapted Physical Activity Quarterly* **28,** 146-156.

Armstrong, N., Williams, J. and Ringham, D. (1988) Peak oxygen uptake and progressive shuttle run performance in boys aged 11-14 years. *British Journal of Physical Education* **19,** 10-11.

Aslan, E., Muniroglu, S., Alemdaroglu, U. and Karakoc, B. (2012) Investigation of the performance responses of yo-yo and shuttle run tests with the treadmill run test in young soccer players. *Pamukkale Journal of Sport Sciences* **3,** 104-112.

Aziz, A.R., Chia, M.Y. and The, K.C. (2005a) Measured maximal oxygen uptake in a multi-stage shuttle test and treadmill-run test in trained athletes. *Journal of Sports Medicine and Physical Fitness* **45,** 306-314.

Aziz, A.R., Mukherjee, S., Chia, M. and The, K.C. (2007) Relationship between measured maximal oxygen uptake and aerobic endurance performance with running repeated sprint ability in young elite soccer players. *Journal of Sports Science and Medicine* **7,** 401-407.

Aziz, A.R., Tan, F.H.Y. and The, K.C. (2005b) A pilot study comparing two field tests with the treadmill run test in soccer players. *Journal of Sports Science and Medicine* **4,** 105-112.

Bandyopadhyay, A. (2011) Validity of the 20 meter multi-stage shuttle run test for estimation of maximum oxygen uptake in male university students. *Indian Journal of Physiology and Pharmacology* **55,** 221-226.

Bandyopadhyay, A. (2013) Validity of the 20 meter multi-stage shuttle run test for estimation of maximum oxygen uptake in female university students. *Indian Journal of Physiology and Pharmacology* **57,** 77-83.

Banerjee, A.K., Chatterjee, P., Majumdar, P. and Chatterjee, P. (2005) Validity of 20-m multi stage shuttle run test for estimation of maximum oxygen uptake in women boxers. *Indian Journal of Physical Education, Sports Medicine and Exercise Sciences* **3–4,** 15-19.

Barnett, A., Chan, L.Y.S. and Bruce, I.C. (1993) A preliminary study of the 20-m multistage shuttle run as a predictor of peak VO_2 in Hong Kong Chinese students. *Pediatric Exercise Science* **5,** 42-50.

Batista, M.B., Cyrino, E.S., Arruda, M., Dourado, A.C., Coelho-E-Silva, M.J., Ohara, D., Romanzini, M. and Ronque, E.R. (2013) Validity of equations for estimating VO_2peak from the 20-m shuttle run test in adolescents aged 11-13 years. *Journal of Strength and Conditioning Research* **27,** 2774-2781.

Begg, C.B. and Mazumdar, M. (1994) Operating characteristics of a rank order correlation for publication bias. *Biometrics* **50,** 1088-1101.

Blair, S.N. (2009) Physical inactivity: The biggest public health problem of the 21st century. *British Journal of Sports Medicine* **43,** 1-2.

Cadenas-Sánchez, C., Alcántara-Moral, F., Sánchez-Delgado, G., Mora-González, J., Martínez-Téllez, B., Herrador-Colmenero, M., Jiménez-Pavón, D., Femia, P., Ruiz, J.R. and Ortega, F.B. (2014) Evaluación de la capacidad cardiorrespiratoria en niños de edad preescolar: Adaptación del test de 20m de ida y vuelta. *Nutrición Hospitalaria* **30,** 1333-1343. (In Spanish)

Castagna, C., Impellizzeri, F.M., Manzi, V. and Ditroilo, M. (2010) The assessment of maximal aerobic power with the multistage fitness test in young women soccer players. *Journal of Strength and Conditioning Research* **24,** 1488-1494.

Castagna, C., Impellizzeri, F.M., Rampinini, E., D'Ottavio, S. and Manzi, V. (2008) The Yo-Yo intermittent recovery test in basketball players. *Journal of Science and Medicine in Sport* **11,** 202-208.

Castro-Piñero, J., Artero, E.G., España-Romero, V., Ortega, F.B., Sjöström, M. and Ruiz, J.R. (2010) Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* **44,** 934-943.

Castro-Piñero, J., Ortega, F.B., Keating, X.D., González-Montesinos, J.L., Sjöstrom, M. and Ruiz, J.R. (2011) Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years; influence of weight status. *Nutrición Hospitalaria* **26,** 572-578.

Chatterjee, P., Banerjee, A.K. and Das, P. (2010a) Applicability of an indirect method to predict maximum oxygen uptake in young badminton players of Nepal. *International Journal of Sports Science & Engineering* **4,** 209-214.

Chatterjee, P., Banerjee, A.K. and Das, P. (2011) A prediction equation to estimate the maximum oxygen uptake of school-age girls from Kolkata, India. *Malaysian Journal of Medical Sciences* **18,** 25–29.

Chatterjee, P., Banerjee, A.K., Das, P. and Dednath, P. (2009) A regression equation to predict VO_2max of young football players of Nepal. *International Journal of Applied Sport Sciences* **21,** 113-121.

Chatterjee, P., Banerjee, A.K., Das, P. and Dednath, P. (2010b) A regression equation for the estimation of maximum oxygen uptake in Nepalese adult females. *Asian Journal of Sports Medicine* **1,** 41-45.

Chatterjee, P., Banerjee, A.K., Das, P. and Debnath, P. (2010c) A regression equation for the estimation of VO2max in Nepalese male adults. *Journal of Human Sport and Exercise* **5,** 127-133.

Chatterjee, P., Banerjee, A.K., Das, P., Debnath, P. and Chatterjee, P. (2005) Validity of the 20-m multi stage shuttle run test for the prediction of maximal oxygen uptake in trainee footballers. *Indian Biologist* **37,** 31-35.

Chatterjee, P., Banerjee, A.K., Das, P., Debnath, P. and Chatterjee, P. (2006a) A regression equation to predict VO2 max of trainee badminton players. *Biomedicine* **26,** 47-52.

Chatterjee, P., Banerjee, A.K., Das, P., Dednath, P. and Chatterjee, P. (2008a) Regression equations to predict VO_2max in untrained boys & junior sprinters of Kolkata. *Journal of Exercise Science & Physiotherapy* **4,** 104-108.

Chatterjee, P., Banerjee, A.K., Das, P., Dednath, P. and Chatterjee, P. (2008b) Validity of 20-m multi stage shuttle run test for prediction of maximum oxygen uptake in Indian female university students. *Kathmandu University Medical Journal* **6,** 176-180.

Chatterjee, P., Banerjee, A.K., Debnath, P., Bas, P. and Chatterjee, B. (2006b) Validity of 20-metre multi stage shuttle run test for estimation of maximum oxygen uptake in Indian male university students. *African Journal for Physical, Health Education, Recreation and Dance* **12,** 461-467.

Chatterjee, P., Banerjee, A.K., Dednath, P., Das, P. and Chatterjee, P. (2008c) A regression equation for the estimation of maximum oxygen uptake in Indian male university students. *International Journal of Sports Science* **20,** 1-9.

Chatterjee, P., Banerjee, A.K. and Majumdar, P. (2006c) Validity of the 20-m multi stage shuttle run test for the prediction of VO_2max in junior taekwondo players of India. *International Journal of Sports Science* **18,** 1–7.

Chatterjee, P., Banerjee, A.K., Majumdar, P. and Chatterjee, P. (2007) A regression equation to predict VO_2max in junior taekwondo players. In: *Proceeding of 8th Annual Conference of West Bengal Association of Sports Medicine, January 20, West Bengal.*

Chatterjee, P. and Das, P. (2013) Applicability of 20-M MST as a predictor of maximal oxygen uptake for use with trainee taekwondo players of Nepal. *Indian Journal of Applied Research* **3,** 2249-2255.

Chia, M., Aziz, A.R., Tan, F. and The, K.C. (2005) Examination of the performance of youth soccer players in a 20-metre shuttle run test and a treadmill run test. *Advances in Exercise and Sports Physiology* **11,** 95-101.

Chen, M.J., Fan, X. and Moe, S.T. (2002) Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: A meta-analysis. *Journal of Sports Sciences* **20,** 873-899.

Cohen, J.A. (1992) Power primer. *Psychological Bulletin* **112,** 155-159.

Cooper, H., Hedges, L.V. and Valentine, J.C. (2009) *The handbook of research synthesis and meta-analysis.* 2nd edition. Sage, New York.

Council of Europe Committee for the Development of Sport. Eurofit (1988) *Handbook for the EUROFIT tests of physical fitness.* Edigraf editoriale grafica, Rome.

Cunningham, L.N., Cama, G., Cilia, G. and Bazzano, O. (1994) Relationship of VO_2max with the 1-mile run and 20 meter shuttle test with youth aged 11 to 14 years. *Medicine & Science in Sports & Exercise* **26,** 8209.

De Souza, A.M., Potts, J.E., Bell, S. and D'Abreo, R. (2010) Using the 20m-shuttle run test to predict VO_2max in female national level field hockey players. *Medicine & Science in Sports & Exercise* **42,** 157.

Dickau, L. (2011) *Examination of aerobic and anaerobic contributions to yo-yo intermittent recovery level 1 test performance in female adolescent soccer players.* Master thesis, University of Victoria, Victoria.

Dong-Ho, P., Jung-Ran, S., Sang-Hyun, L. and Chang-Sun, K. (2014) 20 m 점증 왕복달리기 검사를 이용한 여중생의 VO_2max 추정식 개발. *Exercise Science* **23,** 1-11. (In Korean)

Flather, M.D., Farkouh, M.E., Pogue, J.M. and Yusuf, S. (1997) Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials* **18,** 568-579.

Flouris, A.D., Koutedakis, Y., Nevill, A., Metsios, G.S., Tsiotra, G. and Parasiris, Y. (2004) Enhancing specificity in proxy-design for the assessment of bioenergetics. *Journal of Science and Medicine in Sport* **7,** 197-204.

Flouris, A.D., Metsios, G.S., Famisis, K., Geladas, N. and Koutedakis, Y. (2010) Prediction of VO_2max from a new field test based on portable indirect calorimetry. *Journal of Science and Medicine in Sport* **13,** 70-73.

Flouris, A.D., Metsios, G.S. and Koutedakis, Y. (2006) Contribution of muscular strength in cardiorespiratory fitness tests. *Journal of Sports Medicine and Physical Fitness* **46,** 197–201.

Gadoury, C., Léger, L.A. (1986) Validité de l'épreuve de course navette de 20m avec paliers de 1minute et du physistest canadien pour prédire le VO_2max des adultes. *Revue des Sciences et Techniques des Activités Physiques et Sportives* **7,** 57-68.(In French)

Goosey-Tolfrey, V.L. and Tolfrey, K. (2008) The multi-stage fitness test as a predictor of endurance fitness in wheelchair athletes. *Journal of Sports Sciences* **26,** 511-517.

Green, M.S., Esco, M.R., Martin, T.D., Pritchett, R.C., McHugh, A.N. and Williford, H.N. (2013) Cross-validation of two 20-M shuttle-run tests for predicting VO_2max in female collegiate soccer players. *Journal of Strength and Conditioning Research* **27,** 1520-1528.

Hamlin, M.J., Fraser, M., Lizamore, C.A., Draper, N., Shearman, J.P. and Kimber, N.E. (2014) Measurement of cardiorespiratory fitness in children from two commonly used field tests after accounting for body fatness and maturity. *Journal of Human Kinetics* **27,** 83-92.

Hamlin, M.J., Fraser, M., Lizamore, C.A., Fryer, S., Draper, N., Shearman, J.P. and Kimber, N.E. (2013) Criterion-related validity of the 20-m shuttle and the 550-m distance run in 8-13 year-old children. In: *Proceeding of British Association for Sport and Exercise Science 2013 Conference, September 3-5, Preston.*

Hemmings, S.J., Nevill, M.E. and Nevill, A. (2003) Validation of the 20-m multi-stage shuttle test as a predictor of peak oxygen uptake in young elite sports performers. *Journal of Sports Science* **21,** 277.

Howley, E.T., Bassett, D.R. and Welch, H.G. (1995) Criteria for maximal oxygen uptake: Review and commentary. *Medicine & Science in Sports & Exercise* **27,** 1292-1301.

Hunter, J.E. and Schmidt, F.L. (2004) *Methods of meta-analysis: Correcting error and bias in research findings.* 2nd edition. Sage, Newbury Park.

Kim, Y., Park, I. and Kang, M. (2012) Convergent validity of the International Physical Activity Questionnaire (IPAQ): Meta-analysis. *Public Health Nutrition* **16,** 440-452.

Kloyiam, S., Breen, S., Jakeman, P., Conway, J. and Hutzler, Y. (2011) Soccer-specific endurance and running economy in soccer players with cerebral palsy. *Adapted Physical Activity Quarterly* **28,** 354-367.

Kodama, S., Saito, K., Tanaka, S., Maki, M., Yachi, Y., Asumi, M., Sugawara, A., Totsuka, K., Shimano, H., Ohashi, Y., Yamada, N. and Sone, H. (2009) Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* **301,** 2024-2035.

Kuisis, S.M. (2007) *Comparative validity of ice-skating performance tests to assess aerobic capacity.* Master thesis, University of Pretoria, Pretoria.

LaMontagna, R. (1991) *Validity and reliability of the 20-meter shuttle test in American females 19-34 years of age.* Master thesis, Northern Illinois University, DeKalb.

Léger, L.A. and Lambert, J. (1982) A maximal multistage 20-m shuttle run test to predict VO_2max. *European Journal of Applied Physiology and Occupational Physiology* **49,** 1-12.

Léger, L.A., Lambert, A., Goulet, A., Rowan, C. and Dinelle, Y. (1984) Capacity aerobic des Quebecois de 6 a 17 ans–test Navette de 20 metres avec paliers de 1 minute. *Canadian Journal of Applied Sport Sciences* **9,** 64-69. (In French)

Léger, L.A., Mercier, D., Gadoury, C. and Lambert, J. (1988) The multistage 20 meter shuttle run test for aerobic fitness. *Journal of Sports Sciences* **6,** 93-101.

Light, R.J. and Pillemer, D.B. (1984) *Summing up: The science of reviewing research.* Harvard University Press, Cambridge.

Lightburne, T.J. (2008) Validation of the Progressive Aerobic Cardiovascular Endurance Run (PACER). Test for children 7-13 years old. *Medicine & Science in Sports & Exercise* **40,** S463.

Lipsey, M.W. and Wilson, D.B. (2001) *Practical meta-analysis.* Sage, Newbury Park.

Liu, N.Y., Plowman, S.A. and Looney, M.A. (1992) The reliability and validity of the 20 meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport* **63,** 360-365.

Mahar, M.T., Crotts, D.J., McCammon, M.R. and Rowe, D.A. (2002) Validity of the PWC170 and the PACER tests as measures of aerobic capacity in 12-14 years old girls. *Medicine & Science in Sports & Exercise* **34,** S294.

Mahar, M.T., Guerieri, A.M., Hanna, M. and Kemble, C.D. (2010) Development of a model to estimate aerobic fitness from PACER performance in adolescents. *AAHPERD 2010 National Convention and Exposition, March 16-20, Indianapolis*.

Mahar, M.T., Guerieri, A.M., Hanna, M.S. and Kemble, C.D. (2011) Estimation of aerobic fitness from 20-m multistage shuttle run test performance. *American Journal of Preventive Medicine* **41,** S117-123.

Mahar, M.T., Hanna, M.S., Kemble, C.D., DuBose, K.D. and Cooper, N. (2013) Estimation of aerobic fitness from PACER performance in older adolescents. *Research Quarterly for Exercise and Sport* **84,** A29.

Mahar, M.T., Welk, G.J., Rowe, D.A., Crotts, D.J. and McIver, K.L. (2006) Development and validation of a regression model to estimate VO_2peak from PACER 20-m shuttle run performance. *Journal of Physical Activity and Health* **3,** 34-46.

Mahoney, C. (1992) 20-MST and PWC 170 validity in non-caucasian children in the UK. *British Journal of Sports Medicine* **26,** 45-47.

Mayorga-Vega, D., Merino-Marban, R. and Viciana, J. (2014a) Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine* **13,** 1-14.

Mayorga-Vega, D., Viciana, J., Cocca, A. and Merino-Marban, R. (2014b) Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis. *Journal of Human Sport and Exercise* **9,** 188-200.

Matsuzaka, A., Takahashi, Y., Yamazoe, M., Kumakura, N., Ikeda, A., Wilk, B. and Bar-Or, O. (2004) Validity of the multistage 20-m shuttle-run test for Japanese children, adolescents, and adults. *Pediatric Exercise Science* **16,** 113-125.

McIver, K., Pfeiffer, K.A., Mahar, M.T. and Pate, R.R. (2004) Associations between peak VO_2 and field tests of cardiorespiratory fitness in adolescent males. *Medicine & Science in Sports & Exercise* **36,** 134.

McVeigh, S.K., Payne, A.C. and Scott, S. (1995) The reliability and validity of the 20-meter shuttle test as a predictor of peak oxygen uptake in Edinburgh school children, age 13 to 14 years. *Pediatric Exercise Science* **7,** 69-79.

Metsios, G.S., Flouris, A.D., Koutedakis, Y. and Theodorakis, Y. (2006) The effect of performance feedback on cardiorespiratory fitness field tests. *Journal of Science and Medicine in Sport* **9,** 263-266.

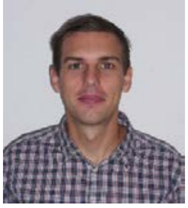
Midgley, A.W., Bentley, D.J., Luttikholt, H., McNaughton, L.R. and Millet, G.P. (2008) Challenging a dogma of exercise physiology. Does an incremen-

tal exercise test for valid VO_2max determination really need to last between 8 and 12 minutes? *Sports Medicine* **38,** 441-447.
Mombiedro, C., Léger, L.A., Cazorla, G.A., Delgado, M., Gutierrez, A., Prost, A. and Roy, J.Y. (1992) Validité du test de course-navette de 20 metres pour predire le VO_2max d'athletes d'endurance. *Science et Motricite* **17,** 3-10. (In French)
O'Gorman, D., Hunter, A., Mc-Donnacha, C., Kirwan, J.P. (2000) Validity of field tests for evaluating endurance capacity in competitive and international level sports participants. *Journal of Strength and Conditioning Research* **14,** 62–67.
Orwin, R.G. (1983) A fail-safe n for effect size in meta-analysis. *Journal of Educational Statistics* **8,** 157-159.
Paliczka, V.J., Nichols, A.K. and Boreham, C.A. (1987) A multi-stage shuttle run as a predictor of running performance and maximal oxygen uptake in adults. *British Journal of Sports Medicine* **21,** 163–165.
Paradisis, G., Zacharogiannis, E., Argeitaki, P. and Smirniotou, A. (2013) Validity of 20mmst for predicting VO_2max. *Medicine & Science in Sports & Exercise* **45,** 687-688.
Paradisis, G.P., Zacharogiannis, E., Mandila, D., Smirtiotou, A., Argeitaki, P. and Cooke, C.B. (2014) Multi-stage 20-m shuttle run fitness test, maximal oxygen uptake and velocity at maximal oxygen uptake. *Journal of Human Kinetics* **41,** 81-87.
Pescatello, L.S., Arena, R., Riebe, D. and Thompson, P.D. (2014) *ACSM's guidelines for exercise testing and prescription.* 9th edition. Wolters Kluwer/ Lippincott Williams & Wilkins, Philadelphia.
Pitetti, K.H., Fernhall, B. and Figoni, S. (2002) Comparing two regression formulas that predict VO_2peak using the 20-m shuttle run for children and adolescents. *Pediatric Exercise Science* **14,** 125-134.
Poortmans, J., Vlaeminck, M., Collin, M. and Delmotte, C. (1986) Estimation indirecte de la puissance aérobie maximale d'une population bruxelloise mascuiine at féminine âgée de 6 a 23 ans. Comparaison avec une technique directe de la mesure de la consommation maximale d'oxygène. *Journal de Physiologie* **81,** 195-201. (In French)
Quinart, S., Mougin, F., Simon-Rigaud, M.L., Nicolet-Guénat, M., Nègree, V. and Regnarda, J. (2014) Evaluation of cardiorespiratory fitness using three field tests in obese adolescents: Validity, sensitivity and prediction of peak VO_2. *Journal of Science and Medicine in Sport* **17,** 521-525.
Ramsbottom, R., Brewer, J. and Williams, C. (1988) A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine* **22,** 141-144.
Riddoch, C.J. (1990) *The Northern Ireland health and fitness survey-1989: The fitness, physical activity, attitudes and lifestyles of Northern Ireland post-primary schoolchildren.* The Queen's University of Belfast, Belfast.
Rowland, T.W. (1993) Does peak VO_2 reflect VO_2max in children?: Evidence from supramaximal testing. *Medicine & Science in Sports & Exercise* **25,** 689-693.
Ruiz, J.R., Castro-Pinero, J., Artero, E.G., Ortega, F.B., Sjostrom, M., Suni, J. and Castillo, M.J. (2009) Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine* **43,** 909-923.
Silva, G., Oliveira, N.L., Aires, L., Mota, J., Oliveira, J. and Ribeiro, J.C. (2012) Calculation and validation of models for estimating VO_2max from the 20-m shuttle run test in children and adolescents. *Archives of Exercise in Health and Disease* **3,** 145-152.
Stickland, M.K., Petersen, S.R. and Bouffard, M. (2003) Prediction of maximal aerobic power from the 20-m multistage shuttle run test. *Canadian Journal of Applied Physiology* **28,** 272-282.
Suminski, R.R., Ryan, N.D., Poston, C.S. and Jackson, A.S. (2004) Measuring aerobic fitness of Hispanic youth 10 to 12 years of age. *International Journal of Sports Medicine* **25,** 61-67.
Thomas, A., Dawson, B. and Goodman, C. (2006) The yo-yo test reliability and association with a 20-m shuttle run and VO_2max. *International Journal of Sports Physiology and Performance* **1,** 137-149.
Terrin, N., Schmid, C.P. and Lau, J. (2005) In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology* **58,** 894-901.
Tsiaras, V., Zafeiridis, A. and Dipia, K. (2010) Prediction of peak oxygen uptake from a maximal treadmill test in 12-to 18-years old active male adolescents. *Pediatric Exercise Science* **22,** 624-637.
Van Mechelen, W., Hlobil, H. and Kemper, H.C. (1986) Validation of two running test as estimates of maximal aerobic power in children. *European Journal of Applied Physiology and Occupational Physiology* **55,** 503-506.
Van Praagh, E., Bedu, M., Falgairette, G., Fellmann, N. and Coudert, J. (1988) Comparaison entre VO_2max direct et indirect chez l'enfant de 7 et 12 ans. Validation d'une épreuve de terrain. *Science & Sports* **3,** 327-332. (In French)
Varness, T., Carrel, A.L., Eickhoff, J.C. and Allen, D.V. (2009) Reliable prediction of insulin resistance by a school-based fitness test in middle-school children. *International Journal of Pediatric Endocrinology* **2009,** 1-7.
Von Haaren, B., Härtel, S., Seidel, I., Schlenker, L. and Bös, K. (2011) Die validität des 6-minuten-laufs und 20m shuttle runs bei 9- bis 11-jährigen kindern. *Deutsche Zeitschrift für Sportmedizin* **62,** 351-355. (In German)

Key points

- Overall the 20-m shuttle run test has a moderate-to-high mean criterion-related validity for estimating cardiorespiratory fitness.
- The criterion-related validity of the 20-m shuttle run test is significantly higher for adults than for children. However, when the performance score is combined with other variables, the criterion-related validity value increases considerably among children.
- Sex and maximum oxygen uptake level of individuals seem not to affect the criterion-related validity of the 20-m shuttle run test.
- When individuals' maximum oxygen uptake attained during a laboratory-based test is not feasible, the 20-m shuttle run test seems to be a useful alternative for estimating cardiorespiratory fitness.

AUTHOR BIOGRAPHY



Daniel MAYORGA-VEGA
Employment
Department of Physical Education and Sport, University of Granada, Spain
Degree
MSc
Research interests
Measurement and evaluation, health-related physical fitness, health-enhancing physical activity, physical education-based interventions, motivation toward physical activity.
E-mail: dmayorgavega@gmail.com

Pablo AGUILAR-SOTO
Employment
Department of Physical Education and Sport, University of Granada, Spain
Degree
MSc
Research interests
Measurement and evaluation, health-related physical fitness, physical education-based interventions.
E-mail: aguilarsotopablo@gmail.com

Jesús VICIANA
Employment
Department of Physical Education and Sport, University of Granada, Spain
Degree
PhD
Research interests
Health-related physical fitness, health-enhancing physical activity, physical education-based interventions, motivation toward physical activity.
E-mail: jviciana@ugr.es

✉ **Daniel Mayorga-Vega**
Department of Physical Education and Sport, University of Granada, Alfacar street, 18011, Granada, Spain

Paper VI

CRITERION-RELATED VALIDITY OF THE DISTANCE- AND TIME-BASED WALK/RUN FIELD TESTS FOR ESTIMATING CARDIORESPIRATORY FITNESS: A SYSTEMATIC REVIEW AND META-ANALYSIS

Mayorga-Vega, D., Bocanegra-Parrilla, R., Ornelas, M., & Viciana, J.

Plos One

Submitted

Criterion-Related Validity of the Distance- and Time-Based Walk/Run Field Tests for Estimating Cardiorespiratory Fitness: A Systematic Review and Meta-Analysis

Criterion-Related Validity of Walk/Run Field Tests

Daniel Mayorga-Vega[1*], Raúl Bocanegra-Parrilla[1], Martha Ornelas[2], Jesús Viciana[1]

[1] Department of Physical Education and Sport, University of Granada, Granada, Spain

[2] Faculty of Sciences of Physical Culture, Autonomous University of Chihuahua, Chihuahua, Mexico

* Corresponding autor

E-mail: dmayorgavega@gmail.com (DMV)

Abstract

Objectives

The main purpose of the present meta-analysis was to examine the criterion-related validity of the distance- and time-based walk/run tests for estimating cardiorespiratory fitness among apparently healthy individuals.

Materials and Methods

Relevant studies were searched from seven electronic bibliographic databases up to August 2015, as well as from other modes of searching. The Hunter-Schmidt's psychometric meta-analysis approach was conducted to estimate the population criterion-related validity of the following walk/run tests: 5,000 m, 3 miles, 2 miles, 3,000 m, 1.5 miles, 1 mile, 1,000 m, ½ mile, 600 m, 600 yd, ¼ mile, 15 min, 12 min, 9 min, and 6 min.

Results

From 123 studies that were included in the present meta-analysis, a total of 200 correlation values were analyzed. The overall results showed that the criterion-related validity of the walk/run tests for estimating maximum oxygen uptake ranged from low to moderate (r_p = 0.42-0.79), being the 1.5 mile (r_p = 0.79, 0.73-0.85) and 12 min walk/run tests (r_p = 0.78, 0.72-0.83) those that showed the higher average criterion-related validity coefficient of distance- and time-based tests, respectively. The present meta-analysis also showed that sex, age and maximum oxygen uptake level do not seem to affect the criterion-related validity of the walk/run tests.

Conclusions

When an individual's maximum oxygen uptake attained during a laboratory-based test is not feasible, the 1.5 mile and 12 min walk/run tests seem to be useful alternatives for estimating cardiorespiratory fitness. Although among children these walk/run tests showed a similar validity as the 20-m shuttle run test, among adults the 20-m shuttle run test should be used instead. As in the application of any physical fitness field test, evaluators must be aware that the performance score of the walk/run field tests is simply an estimation and not a direct measure of cardiorespiratory fitness.

Introduction

Physical fitness constitutes an integrated measure of all the functions and structures involved in the performance of physical activity [1]. Particularly cardiorespiratory fitness reflects the overall capacity of the cardiovascular and respiratory systems to supply oxygen during sustained physical activity, as well as the ability to carry out prolonged exercise [2]. Currently there is strong evidence that cardiorespiratory fitness constitutes an important predictor of morbidity and mortality [3]. Therefore, cardiorespiratory fitness is nowadays considered one of the most powerful markers of health, even above other traditional indicators such as weight status, blood pressure or cholesterol level [4].

Cardiorespiratory fitness testing may help to identify a target population for primary prevention and health promotion policies [5]. Cardiorespiratory fitness is typically operationalized as the maximal oxygen uptake (VO_2max) attained during a graded maximal exercise test [6]. Particularly the VO_2max attained during a laboratory-based and graded maximal exercise test is commonly considered the criterion measure or “gold standard” of cardiorespiratory fitness [6]. Nevertheless, since the direct determination of VO_2max by laboratory testing requires sophisticated and expensive equipment, qualified examiners, and long testing sessions, this technique is not feasible in several settings such as in sports clubs, physical education lessons, or large scale research studies [7].

Unlike the laboratory tests which directly determinate VO_2max, in the above mentioned settings the performance score obtained during cardiorespiratory fitness field tests could be a useful alternative [7]. Cardiorespiratory fitness field tests require minimal equipment, are easy to administer, are not too time-consuming, and allow a large number of individuals to be tested simultaneously. Although many cardiorespiratory fitness field tests have been designed up today, the walk/run field tests are probably the most widely used for estimating cardiorespiratory fitness [8]. The walk/run field tests involve walking and/or running as much as possible during a fixed distance or time, where maintaining a good pace is an important issue in order to obtain a true reflection of actual cardiorespiratory fitness [9]. While the distance-based walk/run tests are conducted over a set distance, and the finishing time is recorded, the time-based walk/run tests are characterized for being conducted during a set time, and the distance covered recorded. Since the early interest in cardiorespiratory fitness testing in the 1950-60s, many distance- and time-based walk/run tests have been proposed [8]. Nevertheless, today there is still no consensus regarding the appropriate length of the walk/run test for estimating cardiorespiratory fitness [10].

Each primary study that is published about criterion-related validity of the distance- and time-based walk/run tests only constitutes a single piece of a constantly growing body of evidence [11]. For example, when Cooper [12] studied the criterion-related validity of the 12 min walk/run test, he found a high correlation coefficient, however, later other authors found a moderate [13] or even low association [14]. To make sense of the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses [11,15,16]. Therefore, meta-analyses remain a useful tool for the evaluation of evidence [11], forming a critical process for the development of theory in science [16].

Recently some meta-analytic studies have examined the criterion-related validity of different widely used physical fitness field tests such as the sit-and-reach test [17], toe-touch test [18], and 20-m shuttle run test [7]. Regarding the distance- and time-based walk/run field tests, Safrit, Hooper, Ehlert, Costa, and Patterson [19] carried out the first meta-analysis on the topic almost 30 years ago. However, from that date many primary research studies have been published. Additionally, some distance- and time-based tests were not taken into account in the authors´ review. Since these authors performed the analysis combining all the tests together, besides the well-known methodological problem of dependency, key issues such as which is the "long enough" distance- or time-based test neither could be addressed. Finally, apart from the sex and age of individuals, the potential moderator effects of other important issues on the criterion-related validity such as the individuals´ fitness levels, adding some other individuals´ characteristics to the performance score, or the measurement unit of the criterion test were not examined.

Unfortunately, to our knowledge there is not any recent meta-analysis examining the criterion-related validity of the distance- and time-based walk/run tests, and there is not any meta-analysis addressing all the above mentioned issues. Examining these questions would help scientists, practitioners and users to select the best feasible and valid test for estimating cardiorespiratory fitness. Consequently, the purposes of the present meta-analysis were: (a) to estimate and compare the overall population mean of the criterion-related validity coefficients of the distance- and time-based walk/run tests for estimating cardiorespiratory fitness among apparently healthy individuals (for only performance score and performance score with other variables); (b) to examine the influence of some study features (sex, age, and level of VO_2max of the participants) on criterion-related validity coefficients of the distance- and time-based walk/run tests (between-study analyses); and (c) to compare the values of the criterion-related validity coefficients between the only performance score and the performance score combined with other variables, as well as between the VO_2max relative to body mass and the

VO_2max absolute, VO_2max relative to fat-free mass and maximal aerobic speed (within-study analyses).

Materials and Methods

All the methodological procedure followed in the present study was based on previous general literature about meta-analyses [11,15,16], as well as specifically in the meta-analyses of the criterion-related validity of physical fitness field tests [7,17,18]. The present manuscript also reproduces some information already reported in detail elsewhere [7,17,18].

Data Sources and Search Strategy

The following seven electronic bibliographic databases were searched through until August 2015: Web of Science^{TM} (all databases), Scopus, SPORTDiscus with Full Text, CINAHL, Cochrane Library, ProQuest Social Sciences Premium Collection, and ProQuest Dissertations & Theses Global. The searches were carried out in the search field type “Title, abstract, and keywords” or equivalent (e.g. “Topic” for the Web of Science^{TM} database). Any publication format including journal papers and grey literature (i.e. doctoral dissertations, master´s theses, conference proceedings, or technical reports) was examined. Additionally, no language or publication date restrictions were imposed.

The search terms used were based on two concepts: concept one included terms for the walk/run field test and concept two included terms related to validity. Due to the large number of terms related to the first concept, from one to four independent searches were carried out for each walk/run field test. Each search was always combined with the same terms related to the validity concept. The truncated root of certain terms was followed by an asterisk to include multiple variants. Additionally, the keywords that consisted of more than one word were enclosed in quotes. Finally, the terms of the same concept were combined together with the Boolean operator “OR” and then the two concepts were combined using the Boolean operator ‘‘AND’’ [11]. See S1 Appendix for all the specific syntaxes used in the present study.

Based on the results of the Boolean-based search, other modes of searching were carried out. The reference lists of all studies (as well as some related study reviews) were manually searched (also called “snowballing”). Additionally, the reference citations and the researcher publications (first authors) in the Web of Science™ and Scopus databases were also examined. Subsequently, the authors for correspondence (if they were not defined, the first author was used) were contacted by email. Finally, the researcher’s personal lists in ResearchGate and Google Scholar of the first authors were screened. For practical reasons, the search was carried out for one researcher.

Study Selection

The selection criteria to identify studies that examined the criterion-related validity of the distance- and time-based walk/run field tests were: (1) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (2) studies with field tests that consisted of walking/jogging, walking/running, only jogging or only running (i.e. but not only walking) as much as possible during a fixed distance (i.e. 5,000 m, 3 miles, 2 miles, 3,000 m, 1.5 mile, 1 mile, 1,000 m, ½ mile, 600 m, 600 yd, and ¼ mile) or time (i.e. 15 min, 12 min, 9 min and 6 min); (3) studies in which for the criterion measure the VO_2max (or VO_2peak –see potentials and limitations section–) was measured in a standardized and laboratory-based graded exercise test to exhaustion; and (4) studies which reported (or could be computed from raw data reported in the study) the Pearson's *r* zero-order correlation coefficient or simple linear regression (R^2) (or multiple linear regression in case of multiple predictors) of performance scores of the field test (or the performance score with other variables) with the measured VO_2max. No publication format, language or date restriction was imposed.

Since the validity of the standard scales for evaluation study quality has been questioned, rejecting certain studies and accepting others for inclusion in a meta-analysis on the basis of a quality score remain a controversial procedure [11]. Therefore, according to Flather, Farkouh, Pogue, and Yusuf [20], in the present meta-analysis the approach followed has been to ensure that the design has not been flawed (e.g. the VO_2max was measured in a standardized and laboratory-based incremental test to exhaustion), and that there has been a complete reporting of relevant outcomes. For a study to be included in this meta-analysis, sample size, protocol of the walk/run field test, unit and protocol of the criterion measure test, statistical test, and value of the criterion-related validity were considered to be critical. In the event that the authors failed to identify any critical study feature and it could not be retrieved, the study was not included in the meta-analysis. Since all the included studies had the same "quality" (i.e. met the four selection criteria), no additional analyses on risk of bias within studies were carried out. Studies selection criteria were examined by two independent researchers. When doubt or disagreement occurred, a consensus was achieved through discussion.

Data Extraction

For the present meta-analysis, from each selected study the following data were coded: Identification number, study reference, type of publication (1 = journal paper, 2 = grey literature –e.g. doctoral dissertation, master´s thesis, conference proceeding, or technical report-), sample size (n), sex of participants (1 = men, 2 = women, 3 = men and women), age of participants (1 = children, < 18 years, 2 = adults, ≥ 18 years, 3 = children and adults), field test (1 = 5,000 m, 2 = 3 miles, 3 = 2 miles, 4 = 3,000 m, 5 = 1.5 miles, 6 = 1 mile, 7 = 1,000 m, 8 = ½ mile, 9 = 600 m, 10 = 600 yd, 11 = ¼ mile, 12 = 15 min, 13 = 12 min, 14 = 9 min, 15 = 6 min), criterion measure protocol (1 = treadmill test, 2 = cycle ergometer test, 3 = other), measurement unit of the criterion test (1 = VO_2max relative to body mass, 2 = VO_2max absolute, 3 = VO_2max relative to fat-free mass, 4 = maximal aerobic speed, 5 = other), mean value of the measurement criterion, standard deviation value of the measurement criterion, reliability of the field test (intraclass correlation coefficient), reliability of the measurement criterion (intraclass correlation coefficient), statistical test used for the criterion-related validity (1 = Pearson's r correlation coefficient, 2 = R^2 simple or multiple linear regression), and criterion-related validity value (separately for only performance score and multiple predictors). In addition, any observations were also registered when a special issue was found. In the event that the authors failed to identify any study feature, they were contacted to retrieve it. If the study feature was not retrieved, the data was omitted. Coding studies were carried out by two independent researchers. When doubt or disagreement occurred, a consensus was achieved through discussion.

Data Analyses

According to Schmidt and Hunter [16], in the present meta-analysis Pearson's zero-order correlation coefficient (r) was considered the unit of measurement as an indication of the criterion-related validity of the walk/run field tests. Therefore, when the validity values were reported as R^2, it was previously transformed by the square root. Although the correlation between the distance-based walk/run tests and the criterion measure was negative, for the present meta-analysis all the correlation coefficients were previously transformed to positive. Finally, the studies carried out with a small sample (defined as less than 10 participants) were not included in the meta-analyses.

Dependency Issues

To avoid dependency issues in the present meta-analysis, an exhaustive examination of the selected studies was carried out. Since the most examined studies used the VO_2max relative to body mass (i.e. expressed in ml/kg/min or similar) as the measurement criterion, the correlation coefficients with this variable were used for the main analyses. When these studies also reported the results of criterion-related validity using additional variables (i.e., the VO_2max absolute, the VO_2max relative to fat-free mass and/or the maximal aerobic speed), these validity coefficients were only used for the within-study analyses to compare with the VO_2max relative to body mass. Since some studies used multiple performance scores of the field tests for examining the criterion-related validity, the average value was used. Nevertheless, when authors reported the results of criterion-related validity from the combination of different multiple predictors, only the best model (i.e. higher coefficient value) was used in the present meta-analysis.

If a single study reported more than one *r* value within the same field test, but from different subsamples, each *r* value from different subsamples was assumed to be independent and included them in a single meta-analysis [15]. When, in the same study, data for men/women or children/adults were expressed both separately and together, only the separate data were selected. However, when in the same study, data for the whole sample and subsamples with respect to sex and age categories were expressed, only the whole sample was used. Similarly, when in the same study, data were expressed for different trials, the average value of the coefficients was selected. However, when in the same study, data were expressed for pre- and post-intervention, only the pre-intervention value was used.

Publication Bias

Besides the search strategy followed and selection criteria to avoid availability bias, an examination of the selected studies was carried out to avoid any potential duplication of information retrieved. Similarities between studies of the same authors, with the same correlation coefficients and/or the same sample size were examined. If some selected studies had full or partial duplicated information, these particular correlation values were not analyzed in the meta-analyses. Furthermore, before computing correlations, several exploratory analyses were also conducted for identifying and assessing the impact of any potential publication bias. Firstly, according to Light and Pillemer [21], the scatter plots of correlation coefficients against sample size for each field test were analyzed. Secondly, with the objective of quantifying the outcomes of the scatter plots, based on Begg and Mazumdar [22], a Spearman's rank order correlation between r values and sample size was calculated. Finally, for assessing the impact of any potential publication bias, a file drawer analysis based on effect size was performed to estimate the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean effect size (r_p) down to the small mean r value [23]. According to Cohen's [24] guidelines, the correlation coefficient was interpreted as small when $r <$ 0.30.

Computation of Correlations

The Hunter-Schmidt's psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of the walk/run field tests [16].

This approach estimates the population correlation correcting the observed correlations due to various artefacts such as sampling error and measurement error. The “bare-bone” mean *r* (r_c), corrected for only sampling error was first calculated by weighting each *r* with the respective sample size when aggregating them into r_c. Then, the corrected mean *r* at the population level (r_p) that was unaffected by both sampling error and measurement error was calculated. Since the reliability coefficients (intraclass correlation coefficients) of the field tests were unavailable in most of the included primary studies, the measurement error was corrected using artifact distributions instead of being corrected individually. On the other hand, the measurement error of the criterion test could not be corrected because the reliability was almost unavailable. Finally, the 95% confidence intervals of r_p (95% CI) were calculated.

Moderator Analyses

In the present meta-analysis, due to the low number of *r* values found, partial hierarchical analyses of moderator variables were carried out. According to Schmidt and Hunter [16], to determine the presence of moderator effects which may affect overall criterion-related validity of the field tests (r_p), three different criteria were simultaneously examined: (a) the 95% credibility interval (95% CV) is relatively large or includes the value zero; (b) the percentage of variance accounted for by statistical artefacts is less than 75% of the observed variance in r_p; and (c) the *Q* homogeneity statistic is statistically significant ($p < 0.05$). If at least one of the three criteria were met, it was concluded that the results could be potentially affected by moderator effects. In the presence of moderator effects, criterion-related validity values of the walk/run field tests were analyzed separately by: (a) sex of participants (i.e. men and women); (b) age of participants (i.e. children and adults); and (c) level of VO_2max (i.e. low average level, < P_{50}, and high average level, ≥ P_{50}) (between-study analyses). Additionally, the criterion-related validity values of the field tests for the only performance score were compared with the criterion-related validity with multiple predictors; and the criterion-related validity values with the VO_2max relative to body mass were compared with the criterion-related validity with VO_2max absolute, VO_2max relative to fat-free mass, and maximal aerobic speed (within-study analyses).

The meta-analyses were performed using the software Hunter and Schmidt Meta-Analysis Programs version 2.0 for Windows (Iowa, 2014). All the others statistical analyses and graphs were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20).

Results

Study Description

Fig. 1 shows a flow diagram of the study selection process. Of the 9,546 bibliographic databases search results, potentially relevant publications were identified and retrieved for a more detailed evaluation. Afterward, based on the studies of the Boolean-based search that met the selection criteria (plus several study reviews that were also used for the reference lists mode), other modes of searching were carried out. From the 547 potentially eligible studies, 159 studies met the selection criteria. However, due to full duplication issues, not reporting the criterion-related validity of the VO_2max relative to body mass and/or carrying out the study with a small sample (i.e. defined as less than 10 participants), only 123 studies were included in the present meta-analysis [10,12-14,25-143]. Finally, from the studies that were included in the present meta-analysis, a total of 200 *r* values across the walk/run field tests were retrieved, being 178 *r* correlation coefficients for the criterion-related validity using the only performance score and 22 *r* correlation coefficients for multiple predictors (i.e. the performance score and other variables: age, sex, biological maturation, heart rate, body mass, body height, body mass index, body circumferences, body fat and/or skinfolds).

Fig. 1. Flow diagram of the study selection process.

Some studies retrieved for a more detailed evaluation were not included because they were carried out with non-healthy participants. Other studies were not selected because the protocol of the field test had a different distance/time, only walking was allowed or the participants had to run in a submaximal intensity instead of as much as possible. Then, some potential studies were not selected because the field test was performed on a treadmill instead of a track or they did not use the measured VO_2max during laboratory-based incremental test to exhaustion as a criterion measure. Other studies that were retrieved for a more detailed evaluation were not selected because only cross-validity or criterion-reference validity was examined. Some potential studies were not selected because they only reported a partial correlation coefficient instead of the zero-order correlation coefficients or simple/multiple linear regression. Since in Ruiz´s et al. [10] study the reported criterion-related validity value for multiple predictors was mistaken and the correct value could not be retrieved, the value for multiple predictors was not included. Finally, the full-text of some potential studies was not available. Since all the included studies had the same quality (i.e. met the four selection criteria), the analyses of the risk of bias within studies were unnecessary.

S1 Table presents a summary of the included studies examining the criterion-related validity of walk/run field tests for estimating cardiorespiratory fitness in apparently healthy individuals. Regarding the criterion-related validity for only performance score, a total of 178 r values across 15 walk/run tests (11 distance- and 4 time-based walk/run tests) were retrieved, ranging from 1 value in the 600 m walk/run test to 34 values in the 1 mile walk/run test (median = 9). Total sample sizes for each field test ranged from 28 in the 600 m walk/run test to 1,856 in the 1 mile walk/run test (median = 367). The individual criterion-related validity correlation coefficients of walk/run tests for estimating cardiorespiratory fitness ranged from 0.03 to 0.99 (median = 0.70). Regarding criterion-related validity for performance with other variables, a total of 22 r values across eight walk/run tests (five distance- and three time-based walk/run tests) were retrieved, ranging from one value in the 2 miles, ½ mile, 600 yd, and 6 min walk/run tests to 6 values in the 1 mile walk/run test (median = 1). Total sample sizes for each field test ranged from 44 in the 2 miles walk/run test to 1,156 in the 1 mile walk/run test (median = 87). The individual criterion-related validity correlation coefficients of walk/run tests for estimating cardiorespiratory fitness ranged from 0.65 to 0.99 (median = 0.81).

Publication Bias

Firstly, several exploratory analyses were followed to avoid full or partial duplicated information availability bias. Although 16 research studies met the selection criteria, the correlation coefficient value was not analyzed in the present meta-analysis. On one hand, some doctoral dissertations, master´s theses and/or conference papers were not included because they were published later in a journal paper [e.g.144-146]. On the other hand, from the Cureton´s et al. [51] study the correlation coefficient with the only performance score was not included because the data came from the sum of some samples that had been reported in other journal papers [49,50,96]. However, since these papers did not report the correlation coefficient with multiple predictors, the correlation coefficient with multiple predictors for the overall results and both the only performance score and multiple predictors for the within-study analysis (i.e. for the number of predictors) of the study of Cureton et al. [51] were used.

Exploratory analyses were conducted to identify the presence of potential publication bias. Because the sum of the *r* values for some tests was very small, the following analyses were calculated only for the tests with a *K* equal to 10 or more. Figs. 2 and 3 show the scatter plots of sample size against criterion-related validity coefficients for estimating VO_2max for distance- and time-based walk/run tests, respectively. According to this graphical method, the figures suggest that there was not publication bias for the distance-based walk/run tests. For the time-based walk/run tests explored (i.e. 12 and 9 min walk/run tests), however, the scatter plot suggests the presence of publication bias, because of the absence of *r* values in the lower left hand corner.

Fig. 2. Scatter plot of sample size against criterion-related validity coefficients (*r*) of distance-based walk/run tests for estimating maximal oxygen uptake: (a) 5,000 m walk/run test; (b) 2 miles walk/run test; (c) 3,000 m walk/run test; (d) 1.5 mile walk/run test; and (e) 1 mile walk/run test. Dashed line represents median values of validity coefficients.

Fig. 3. Scatter plot of sample size against criterion-related validity coefficients (*r*) of time-based walk/run tests for estimating maximal oxygen uptake: (a) 12 min walk/run test; and (b) 9 min walk/run test. Dashed line represents median values of validity coefficients.

Similarly to the graphical method, while the results of Spearman's rank order correlation between r values and sample size did not show any statistically significant correlation for the distance-based walk/run tests (5,000 m walk/run test: r = -0.24, p = 0.320; 2 miles walk/run test: r = 0.02, p = 0.940; 3,000 m walk/run test: r = 0.08, p = 0.719; 1.5 miles walk/run test: r = 0.24, p = 0.341; 1 mile walk/run test: r = -0.20, p = 0.255), a statistically significant correlation was found for the 9 min walk/run test (r = -0.68, p = 0.021). However, for the 12 min walk/run test a statistically significant correlation was not found (r = 0.02, p = 0.933). Due to the small number of rs found for most of the tests, the results of both the scatter plot and the Spearman correlation must be interpreted with caution [11,22]. Additionally, empirical evaluations of the funnel plots suggest that their interpretation can be limited [147].

Finally, file drawer analyses based on effect size were carried out for assessing the impact of any potential publication bias. The results of the file drawer analyses are based on effect size for estimating the number of unlocated studies averaging null results (r = 0) that would have to exist to bring the mean r_p down to 0.29. These results are shown in the following lines (in parenthesis the unlocated/located percentage): 5,000 m walk/run test 28 (147%), 3 miles walk/run test 6 (100%), 2 miles walk/run test 19 (136%), 3,000 m walk/run test 16 (133%), 1.5 miles walk/run test 30 (167%), 1 mile walk/run test 36 (106%), 1,000 m walk/run test 3 (150%), 600 m walk/run test 1 (100%), 600 yd walk/run test 8 (100%), ¼ mile walk/run test 5 (63%), 15 min walk/run test 1 (50%), 12 min walk/run test 42 (162%), 9 min walk/run test 14 (127%), and 6 min walk/run test 7 (88%). Except for some tests (i.e. 1,000 m, ¼ mile, and 15 min walk/run tests), the results showed an unlikely number of "lost" studies for most of the examined tests, especially if the percentage of unlocated/located studies is considered (73-167%).

Criterion-Related Validity

Table 1 reports the number of r values studied (K), the total sample size accumulated (N), the overall weighted mean of r corrected for sampling error only (r_c), the overall weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI for overall criterion-related validity correlation coefficients (r_p) for estimating VO_2max across each walk/run field test. Additionally, to detect the presence of moderator effects which may affect overall criterion-related validity of the field tests, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the Q homogeneity statistic were calculated.

Table 1. Results of meta-analyses for overall criterion-related validity correlation coefficients across the distance- and time-based walk/run field tests

Protocols	K	N	r_c	r_p	95% CI[a]	95% CV[b]	% variance[c]	Q statistic
Only performance score								
5,000 m	19	536	0.70	0.73	0.65-0.81	0.44-1.00	31.11	43.96*
3 miles	6	177	0.58	0.60	0.48-0.72	0.45-0.75	73.18	2.30
2 miles	14	951	0.66	0.69	0.62-0.75	0.50-0.88	34.72	27.51*
3,000 m	12	252	0.67	0.70	0.63-0.76	0.70-0.70	100.00	0.00
1.5 mile	18	873	0.76	0.79	0.73-0.85	0.57-1.00	24.74	57.21*
1 mile	34	1,856	0.60	0.62	0.56-0.67	0.36-0.88	32.05	75.32*
1,000 m	2	71	0.72	0.74	0.46-1.00	0.38-1.00	17.25	10.02*
½ mile	9	241	0.53	0.55	0.45-0.65	0.46-0.63	91.49	0.88
600 m	1	28	0.54	0.56	-	-	-	-
600 yd	8	415	0.58	0.60	0.50-0.70	0.39-0.81	44.65	10.36

¼ mile	8	232	0.44	0.45	0.26-0.64	0.02-0.89	33.22	16.80*
15 min	2	120	0.41	0.42	0.21-0.64	0.22-0.62	54.24	1.76
12 min	26	1,204	0.75	0.78	0.72-0.83	0.54-1.00	22.56	93.24*
9 min	11	425	0.63	0.66	0.57-0.74	0.46-0.86	49.16	11.89
6 min	8	367	0.54	0.55	0.49-0.62	0.55-0.55	100.00	0.00
Performance score with other variables								
2 miles	1	44	0.94	0.97	-	-	-	-
1.5 mile	4	210	0.87	0.90	0.86-0.95	0.84-0.96	55.72	3.32
1 mile	6	1,156	0.76	0.79	0.74-0.84	0.68-0.89	26.13	17.73*
½ mile	1	47	0.66	0.69	-	-	-	-
600 yd	1	53	0.65	0.68	-	-	-	-
12 min	3	169	0.78	0.81	0.73-0.89	0.72-0.90	56.65	2.40
9 min	5	283	0.76	0.79	0.74-0.84	0.79-0.79	100.00	0.00
6 min	1	87	0.77	0.80	-	-	-	-

Note. *K*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error of the field tests; [a]95% confidence interval; [b]95% credibility interval; [c]Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of the field tests. * $p < 0.05$

The overall results showed that the criterion-related validity of the distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate and no 95% CI included the value zero. The results of the present meta-analysis also showed that the criterion-related validity of the 1.5 mile and 12 min walk/run tests was statistically significantly higher than the 3 miles, 1 mile, ½ mile, 600 yd, ¼ mile, 15 min, and 6 min walk/run tests ($p < 0.05$). The overall results also showed that the criterion-related validity of the 5,000 m walk/run test was statistically significantly greater than the ¼ mile, 15 min and 6 min walk/run tests ($p < 0.05$). Additionally, the 2 miles and 3,000 m walk/run tests showed a statistically significant higher mean *r* than the 6 min walk/run test ($p < 0.05$). For the other comparisons with the only performance score statistically significant differences were not found ($p > 0.05$). For most of the tests the 95% CV was relatively large, the percentage of variance accounted for by statistical artefacts was less than 75% and/or the *Q* homogeneity statistic was statistically significant ($p < 0.05$). Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study. However, due to the small number of *r*s, moderator analyses were not conducted for the 1,000 m, 600 m and 15 min walk/run tests.

Regarding the multiple predictors, the overall results showed that when the performance score of the walk/run field tests was combined with other variables the mean correlation coefficients of criterion-related validity for estimating VO_2max were moderate-to-very-high. Additionally, when the 95% CI could be calculated, the value zero was not included. The results of the present meta-analysis also showed that the criterion-related validity of the 1.5 mile walk/run test with multiple predictors was statistically significantly higher than the 1 mile and 9 min walk/run tests ($p < 0.05$). However, statistically significant differences between the criterion-related validity of the 1.5 mile and 12 min walk/run tests were not found ($p > 0.05$), as well as neither between the 1 mile, 12 min, and 9 min walk/run tests ($p > 0.05$). Although at least one criterion was met in three of the four tests, due to the small *K* the between-study moderator analyses were not conducted for multiple predictors. Nevertheless, due to the fact that 16 studies reported correlation coefficients of criterion-related validity for both only performance score and combined with other variables, the within-study analysis was conducted as it was hypothesized in the present study. Additionally, the criterion-related validity values with the VO_2max relative to body mass could be compared with the criterion-related validity with VO_2max absolute, VO_2max relative to fat-free mass, and maximal aerobic speed (see moderator analyses).

Moderator Analyses

Table 2 shows the results of between-study moderator analyses to examine the effects of sex (i.e. men and women), age (i.e. children and adults), and level of VO_2max (i.e. low average level, $< P_{50}$ and high average level, $\geq P_{50}$) of participants on criterion-related validity correlation coefficients for estimating VO_2max for each distance- and time-based walk/run field test.

Table 2. Results of the between-study moderator analyses for criterion-related validity correlation coefficients across the distance- and time-based walk/run field tests†

Moderator	Effect	*K*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% variance[c]	*Q* statistic
Sex of participants									
5,000 m	Men	10	302	0.67	0.69	0.56-0.82	0.34-1.00	26.49	28.99*
	Women	5	139	0.80	0.83	0.80-0.85	0.83-0.83	100.00	0.00
3 miles	Men	5	162	0.58	0.60	0.46-0.73	0.41-0.78	61.64	3.25
	Women	1	15	0.66	0.68	-	-	-	-
2 miles	Men	9	816	0.64	0.66	0.59-0.73	0.50-0.82	39.22	14.58
	Women	4	109	0.81	0.84	0.73-0.95	0.67-1.00	39.57	6.38

3,000 m	Men	9	196	0.67	0.69	0.63-0.76	0.69-0.69	100.00	0.00
	Women	3	56	0.70	0.72	0.55-0.89	0.55-0.89	67.71	1.50
1.5 mile	Men	12	585	0.75	0.78	0.70-0.86	0.54-1.00	22.22	43.90*
	Women	4	150	0.74	0.77	0.66-0.88	0.60-0.94	44.33	5.25
1 mile	Men	15	591	0.58	0.60	0.50-0.69	0.31-0.88	36.52	27.24*
	Women	10	415	0.55	0.57	0.43-0.70	0.20-0.93	26.60	28.83*
½ mile	Men	5	125	0.55	0.57	0.42-0.71	0.44-0.69	84.51	0.96
	Women	2	30	0.45	0.46	0.09-0.84	0.16-0.76	66.98	1.03
600 yd	Men	4	158	0.60	0.62	0.50-0.73	0.53-0.71	83.70	0.81
	Women	2	46	0.30	0.31	-0.01-0.63	0.09-0.54	75.30	0.69
¼ mile	Men	3	60	0.24	0.25	0.10-0.40	0.25-0.25	100.00	0.00
	Women	-	-	-	-	-	-	-	-
12 min	Men	13	761	0.76	0.79	0.71-0.87	0.51-1.00	13.75	85.19*
	Women	8	285	0.75	0.78	0.70-0.85	0.64-0.91	54.47	6.99
9 min	Men	6	176	0.69	0.72	0.62-0.81	0.59-0.85	69.83	2.71
	Women	4	165	0.61	0.63	0.46-0.80	0.36-0.90	34.87	7.81*
6 min	Men	3	126	0.53	0.54	0.44-0.64	0.54-0.54	100.00	0.00
	Women	2	55	0.49	0.50	0.39-0.62	0.50-0.50	100.00	0.00
Age of participants									
5,000 m	Children	1	12	0.65	0.67	-	-	-	-
	Adults	18	524	0.71	0.73	0.65-0.81	0.43-1.00	29.27	45.44*
3 miles	Children	-	-	-	-	-	-	-	-
	Adults	5	150	0.56	0.58	0.44-0.72	0.41-0.75	70.26	2.21
2 miles	Children	-	-	-	-	-	-	-	-

	Adults	13	924	0.67	0.69	0.62-0.75	0.49-0.88	32.21	28.58*
3,000 m	Children	2	32	0.65	0.67	0.54-0.80	0.67-0.67	100.00	0.00
	Adults	10	220	0.68	0.70	0.63-0.77	0.70-0.70	100.00	0.00
1.5 mile	Children	3	234	0.73	0.76	0.73-0.78	0.76-0.76	100.00	0.00
	Adults	15	639	0.77	0.80	0.72-0.87	0.54-1.00	19.71	63.85*
1 mile	Children	22	1,499	0.58	0.60	0.53-0.66	0.34-0.85	28.99	56.31*
	Adults	11	330	0.68	0.71	0.61-0.80	0.47-0.94	41.74	16.05
½ mile	Children	5	143	0.47	0.48	0.37-0.60	0.48-0.48	100.00	0.00
	Adults	4	98	0.62	0.64	0.51-0.77	0.61-0.67	98.66	0.06
600 yd	Children	7	371	0.56	0.57	0.47-0.68	0.39-0.76	52.91	6.51
	Adults	1	44	0.78	0.81	-	-	-	-
¼ mile	Children	2	86	0.55	0.57	0.45-0.69	0.57-0.57	100.00	0.00
	Adults	6	146	0.37	0.38	0.13-0.64	-0.13-0.89	33.36	12.52*
12 min	Children	8	246	0.69	0.71	0.67-0.76	0.71-0.71	100.00	0.00
	Adults	17	767	0.76	0.79	0.71-0.87	0.49-1.00	14.99	100.78*
9 min	Children	10	407	0.63	0.65	0.56-0.74	0.45-0.85	48.19	11.24
	Adults	1	18	0.83	0.85	-	-	-	-
6 min	Children	7	337	0.52	0.53	0.48-0.59	0.53-0.53	100.00	0.00
	Adults	1	30	0.79	0.81	-	-	-	-
Level of VO_2max									
5,000 m	Low	9	241	0.74	0.76	0.69-0.83	0.68-0.85	82.41	2.01
	High	10	295	0.68	0.70	0.56-0.84	0.32-1.00	22.76	35.46*
3 miles	Low	2	50	0.64	0.66	0.34-0.98	0.29-1.00	30.32	4.80*
	High	3	92	0.61	0.63	0.57-0.68	0.63-0.63	100.00	0.00

2 miles	Low	6	562	0.69	0.71	0.63-0.80	0.55-0.88	30.46	14.31*
	High	7	354	0.65	0.67	0.57-0.76	0.48-0.85	45.82	8.65
3,000 m	Low	6	115	0.67	0.69	0.60-0.79	0.69-0.69	100.00	0.00
	High	6	137	0.68	0.70	0.61-0.79	0.70-0.70	100.00	0.00
1.5 mile	Low	8	556	0.79	0.81	0.72-0.90	0.58-1.00	13.93	51.66*
	High	9	293	0.71	0.74	0.66-0.81	0.62-0.85	69.71	4.09
1 mile	Low	16	899	0.63	0.65	0.57-0.73	0.37-0.93	25.67	48.41*
	High	16	902	0.58	0.60	0.53-0.67	0.40-0.80	44.16	21.14
½ mile	Low	4	70	0.54	0.55	0.36-0.75	0.38-0.73	80.63	1.00
	High	5	171	0.53	0.54	0.43-0.66	0.54-0.54	100.00	0.00
600 yd	Low	4	257	0.56	0.58	0.43-0.74	0.33-0.84	31.75	8.99*
	High	4	158	0.60	0.62	0.50-0.73	0.53-0.71	83.70	0.81
¼ mile	Low	3	97	0.50	0.51	0.31-0.71	0.31-0.72	62.95	1.85
	High	3	69	0.28	0.29	-0.02-0.60	-0.06-0.64	56.13	2.45
12 min	Low	8	246	0.69	0.71	0.67-0.76	0.71-0.71	100.00	0.00
	High	12	629	0.73	0.75	0.66-0.84	0.48-1.00	19.18	52.83*
9 min	Low	5	213	0.54	0.56	0.48-0.64	0.56-0.56	100.00	0.00
	High	6	212	0.73	0.75	0.66-0.85	0.59-0.92	48.15	6.75
6 min	Low	4	227	0.54	0.55	0.46-0.65	0.51-0.60	95.24	0.21
	High	4	140	0.54	0.55	0.47-0.64	0.55-0.55	100.00	0.00

Note. *K*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error of the field tests; VO_2max, maximal oxygen uptake; [a]95% confidence interval; [b]95% credibility interval; [c]Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of the field tests. †Because some studies mixed categories or data were missing, the overall *K* for some categories is lower for some field tests. * $p < 0.05$

Sex of Participants

The results showed that the criterion-related validity of the analyzed distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate for both men and women. Additionally, no 95% CI included the value zero, except for the 600 yd walk/run test in women. Moreover, among each walk/run field test the 95% CI of mean correlation coefficients overlapped between men and women. As regards the men, the results of the present meta-analysis showed that the criterion-related validity of the 1.5 mile and 12 min walk/run tests was statistically significantly higher than the 1 mile, ¼ mile, and 6 min walk/run tests ($p < 0.05$). Additionally, the ¼ mile walk/run test had a criterion-related validity value statistically significantly lower than all the other tests ($p < 0.05$). Regarding the women, the results showed that the criterion-related validity of the 5,000 m, 2 miles, and 12 min walk/run tests was statistically significantly higher than the 1 mile, 600 yd, and 6 min walk/run tests ($p < 0.05$). Moreover, the 1.5 mile walk/run tests showed a criterion-related validity value statistically significantly greater than the 600 yd and 6 min walk/run tests ($p < 0.05$). On the other hand, it must be pointed out that, according to moderator analysis criteria, at least one criterion was met in most of the walk/run field tests, indicating that the criterion-related validity of these tests separately for sex was still heterogeneous. Finally, because some studies grouped men and women together or the sex values were missing, in Table 2 overall *K* for the sex of participants is lower.

Age of Participants

The results of the present meta-analysis showed that the criterion-related validity of the analyzed distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate for both children and adults in which the 95% CIs did not include the value zero. Moreover, among each walk/run field test the 95% CI of mean correlation coefficients overlapped between children and adults. As regards the children, the results of the present meta-analysis showed that the criterion-related validity of the 1.5 mile walk/run tests was statistically significantly higher than the 1 mile, ½ mile, 600 yd, ¼ mile, and 6 min walk/run tests ($p < 0.05$). Additionally, the 12 min walk/run test had a criterion-related validity value statistically significantly greater than the 1 mile, ½ mile, 6 min walk/run tests ($p < 0.05$). Regarding the adults, the results showed that the criterion-related validity of the 1.5 mile walk/run test was statistically significantly higher than the 3 mile and ¼ mile walk/run tests ($p < 0.05$). Moreover, the 5,000 m and 12 min walk/run tests showed a criterion-related validity value statistically significantly greater than the ¼ mile walk/run test ($p < 0.05$). On the other hand, according to moderator analysis criteria, at least one of the three criteria was met in most of the walk/run field tests, indicating that the criterion-related validity of these tests separately for age was still heterogeneous. Finally, because some studies grouped children and adults together or the age values were missing, in Table 2 overall *K* for the age of participants is lower.

Level of VO_2max

The results showed that the criterion-related validity of the analyzed distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate for both participants with low and high level of VO_2max. Additionally, no 95% CI included the value zero (except for the ¼ mile walk/run test for participants with high level of

VO_2max). Moreover, in each walk/run field test the 95% CI of mean correlation coefficients overlapped between participants with low and high level of VO_2max, except for the 9 min walk/run test in which the participants with higher level showed statistically significant greater values than those with lower level ($p < 0.05$). As regards the participants with low level of VO_2max, the results of the present meta-analysis showed that the criterion-related validity of the 1.5 mile walk/run test was statistically significantly higher than the ¼ mile, 9 min, and 6 min walk/run tests ($p < 0.05$). Additionally, the 5,000 m and 12 min walk/run tests had a criterion-related validity value statistically significantly greater than the 9 min and 6 min walk/run tests ($p < 0.05$). Regarding the participants with high level of VO_2max, the results showed that the criterion-related validity of the 1.5 mile, 12 min, and 9 min walk/run tests was statistically significantly higher than the ½ mile, ¼ mile, and 6 min walk/run tests ($p < 0.05$). Moreover, the 3,000 m walk/run test showed a criterion-related validity value statistically significantly greater than the ¼ mile walk/run test ($p < 0.05$). On the other hand, it must be pointed out that, according to moderator analysis criteria, at least one criterion was met in most of the walk/run field tests, indicating that the criterion-related validity of these tests separately for level of VO_2max was still heterogeneous. Finally, because some studies failed to identify the level of VO_2max, in Table 2 overall *K* for the level of VO_2max is lower.

Table 3 shows the results of within-study moderator analyses to examine the effects of number of predictors (i.e. only performance score and multiple predictors) and the unit of the criterion measure (i.e. VO_2max relative to body mass compared with the VO_2max absolute, VO_2max relative to fat-free mass, and maximal aerobic speed) on criterion-related validity correlation coefficients for estimating VO_2max for each distance- and time-based walk/run field test. Because of the low number of *r*s correlation coefficients for each field test, the within-study analyses were carried out with all the field tests together.

Table 3. Results of the within-study moderator analyses for criterion-related validity correlation coefficients across the distance- and time-based walk/run field tests†

Effect	*K*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% variance[c]	*Q* statistic
Number of predictors[d]								
One predictor	16	1,400	0.63	0.65	0.59-0.71	0.46-0.84	32.45	34.80*
Few predictors	16	1,400	0.76	0.79	0.75-0.82	0.67-0.90	37.44	27.93*
VO_2max absolute								
VO_2max (ml/kg/min)	32	1,809	0.61	0.63	0.59-0.68	0.46-0.81	47.78	36.55
VO_2max (l/min)	32	1,809	0.36	0.37	0.32-0.42	0.25-0.49	80.59	8.05
VO_2max relative to fat-free mass								
VO_2max (ml/kg/min)	16	864	0.67	0.70	0.65-0.74	0.63-0.76	85.45	2.85
VO_2max (ml/kg FFM/min)	16	864	0.48	0.50	0.42-0.58	0.27-0.73	46.49	19.24
Maximal aerobic speed								
VO_2max (ml/kg/min)	11	176	0.57	0.59	0.47-0.70	0.43-0.74	84.13	2.17
MAS (km/h)	11	176	0.68	0.71	0.60-0.81	0.51-0.90	67.30	5.59

Note. *K*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error of the field tests; [a]95% confidence interval; [b]95% credibility interval; [c]Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of the field tests; [d]Only performance score (“one”) or performance score plus other variables (“few”); VO_2max, maximal oxygen uptake; FFM, fat-free mass; MAS, maximal aerobic speed. †Because of the low number of *r*s in each field test, the overall results are reported. * $p < 0.05$

Number of Predictors

The results showed that for the analyzed distance- and time-based tests the only performance score had a moderate mean correlation coefficient of criterion-related validity for estimating VO_2max, but when other variables were added to the performance score the criterion-related validity was moderate-to-high. Additionally, no 95% CI included the value zero. Moreover, the results of the present meta-analysis showed that the criterion-related validity of the only performance score was statistically significantly lower than when other variables were added to the performance score ($p < 0.05$). Nevertheless, it must be pointed out that, according to moderator analysis criteria, at least two of the three criteria were met in both categories, indicating that the criterion-related validity was heterogeneous. The fact that the different walk/run field tests were put together must be also taken into account.

Measurement Unit of the Criterion Test

The results showed that the criterion-related validity of the analyzed distance- and time-based walk/run field tests with the VO_2max relative to body mass and maximal aerobic speed as the measurement unit were moderate, but when the VO_2max absolute and relative to fat-free mass was used instead the criterion-related validity was low. However, no 95% CI included the value zero. The results of the present meta-analysis showed that the criterion-related validity of the walk/run field tests with the VO_2max relative to body mass as the measurement unit was statistically significantly higher than when the VO_2max absolute or relative to fat-free mass was used ($p < 0.05$). However, statistically significant differences between the VO_2max relative to body mass and maximal aerobic speed were not found ($p > 0.05$). Nevertheless, it must be pointed out that, according to moderator analysis criteria, at least one criterion was met in each subcategory, indicating that the criterion-related validity was heterogeneous. Again the fact that the different walk/run field tests were put together must be also taken into account.

Discussion

The choice of a cardiorespiratory fitness test must be based on its feasibility and validity [7]. Although the VO_2max determined directly during a laboratory-based incremental exercise test to exhaustion is considered the criterion measure to assess cardiorespiratory fitness, due to several practical reasons it has the disadvantage of having a limited use in several settings [6]. In settings such as sports clubs, physical education lessons or large scale research studies, since the distance- and time-based walk/run field tests have the advantage of allowing for the evaluation of a large number of people in a short amount of time and with minimal skill and instrumentation, potentially they could be a useful alternative to estimate cardiorespiratory fitness [9].

The first purpose of the present meta-analysis was to estimate and compare the overall population mean of the criterion-related validity coefficients of the distance- and time-based walk/run tests for estimating cardiorespiratory fitness. The overall results showed that the criterion-related validity of the distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate. Although up today many distance- and time-based walk/run tests have been proposed [8], according to the results of the present meta-analysis, the 1.5 mile and 12 min walk/run tests showed a greater average criterion-related validity coefficient. Apart from being statistically significantly higher than most of the shorter distance- and time-based tests (i.e. 1 mile, ½ mile, 600 yd, ¼ mile, and 6 min walk/run tests), these tests were also statistically significantly higher than other tests with a longer length (i.e. 3 miles and 15 min walk/run tests) or simply there were not differences when individuals performed a longer distance/time. Similarly, when multiple predictors were considered, the results of the present meta-analysis showed that the criterion-related validity of the 1.5 mile walk/run test was statistically significantly higher than the 1 mile and 9 min walk/run tests, as well as there were no differences with the 12 min walk/run test.

Therefore, if our purpose is to assess cardiorespiratory fitness through a walk/run field test, the 1.5 mile and 12 min walk/run tests seem to be the best option of distance- and time-based tests, respectively. Additionally, the overall criterion-related validity of these tests has shown to be similar to other cardiorespiratory fitness tests such as the 20-m shuttle run test [7]. It must be highlighted that in the present meta-analysis the overall criterion-related validity of most distance- and time-based walk/run tests was heterogeneous, as well as the number of *K* for many tests was very low (e.g. for only performance score seven of 14 field tests, 47%, had a *K* lower than 10 and for multiple predictors all the tests had a *K* lower than 10). Additionally, not many primary studies comparing the criterion-related validity of various lengths of the walk/run tests were found. Therefore, we should be cautious with the overall results of the present meta-analysis.

The second purpose of this meta-analysis was to examine the influence of some potential moderator factors such as sex, age, and level of VO_2max of the participants on criterion-related validity of the distance- and time-based walk/run tests. According to the findings of the present meta-analysis, sex, age, and fitness levels of individuals do not seem to affect the criterion-related validity of distance- and time-based walk/run tests (exceptionally, for the 9 min walk/run test participants with higher fitness levels showed statistically significant greater values than those with lower levels). Therefore, since the criterion-related validity for men-women, children-adults and low-high level of VO_2max subgroups was similar, the distance- and time-based walk/run tests can be used interchangeably for any subcategory.

Similarly to the findings of the present study, recently Mayorga-Vega et al. [7], carrying out a meta-analytic study about the criterion-related validity of the 20-m shuttle run test, found that sex and fitness levels of individuals did not seem to affect the validity of the test. However, on the contrary to the results of the present meta-analysis, the above mentioned authors found out that the criterion-related validity of the 20-m shuttle run test with the Léger´s protocol was statistically significantly higher among adults than among children. Although among children the 1.5 mile and 12 min walk/run tests showed a similar validity than the 20-m shuttle run test, among adults the 20-m shuttle run test was statistically significantly higher. Therefore, among adults the 20-m shuttle run tests should be used instead the walk/run field tests. A potential reason for these differences could be inherent to the protocols of the field tests. Meanwhile in the walk/run tests individuals have to run as much as possible maintaining a self-pace, the 20-m shuttle run test is characterized to have a rigid standardized protocol where individuals cannot choose their own pace. Particularly, it has been suggested that the starting speed of the 20-m shuttle run could be too high for children [148]. In this line, current evidence suggests that to elicit valid VO_2max values, continuous incremental tests should last at least five minutes [149]. However, Castro-Piñero et al. [150] in a population-based study carried out using the 20-m shuttle run test with Léger's protocol (i.e. starting speed at 8.5 km/h) found that most children lasted less than five minutes. Therefore, meanwhile with the walk/run tests both children and adults can adjust the running pace to their own possibilities, the most widely used protocols of the 20-m shuttle run test such as the Léger [151], Eurofit [152] or QUB [153] could be too high for children. In this line, recent studies have proposed modifications of the 20-m shuttle run test for children with a drastically reduced starting speed (e.g. 4 km/h [107]; 6.5 km/h [148]). Future studies should compare the criterion-related validity of 1.5 mile and/or 12 min walk/run field tests and a modified version of the 20-m shuttle run test with a lower starting speed among children.

For both men-women, children-adults and low-high level of VO_2max subgroups, the 1.5 mile and 12 min walk/run tests seem to be the most appropriate distance- and time-based walk/run tests, respectively. For instance, although other longer distance-based tests showed similar criterion-related validity results to the above mentioned tests, performing longer distance seems to be an unnecessary extra time and effort. On the contrary, due to the lower criterion-related validity values found for estimating the cardiorespiratory fitness, the use of shorter distance/time walk/run tests should be avoided. Surprisingly the 1 mile walk/run test, followed by the ½ mile and ¼ mile walk/run tests, is the cardiorespiratory fitness test more often proposed by the field-based physical fitness batteries for children [8]. On the other hand, the 1.5 mile walk/run test, 12 min walk/run test and the 20-m shuttle run test are proposed by less field-based physical fitness batteries [8]. As regards the FITNESSGRAM® test battery, its test administration manual proposes indistinctly the use of either the 20-m shuttle run test or 1 mile walk/run test [9]. However, according to the results of the present meta-analysis, among children the 1.5 mile walk/run test should be recommended instead as a valid alternative to the 20-m shuttle run tests.

Regarding the third purpose, the present meta-analysis aimed to compare the values of the criterion-related validity coefficients between the only performance score and the performance score combined with other variables, as well as between the VO_2max relative to body mass and the VO_2max absolute, VO_2max relative to fat-free mass and maximal aerobic speed. When multiple predictors were used, the average correlation coefficient was statistically significantly higher than for the only performance score ($r\Delta$ = 0.14). Therefore, apart from the running performance score, adding other individuals´ variables such as age, sex, body mass, body height, body mass index, body circumferences, body fat, and/or skinfolds significantly improves the estimation of the VO_2max. Due to the low number of correlations coefficient found, the fact that these analyses were carried out with different field tests must be taken into account. Similarly to the results of our study, Mayorga-Vega et al. [7] found that for the 20-m shuttle run test with multiple predictors the average correlation coefficient was considerably higher than for the only performance score ($r\Delta$ = 0.11). However, probably because of the low number of correlations that these authors found, this difference was not statistically significant. Another potential reason for these differences could be due to the fact that the validity value of the walk/run tests was lower than the 20-m shuttle run test and, therefore, the change to increase the explained variance was greater.

Finally, the results of the present meta-analysis showed that the criterion-related validity of the walk/run field tests with the VO_2max relative to body mass as the measurement unit was statistically significantly higher than when the VO_2max absolute or relative to fat-free mass was used. Cardiorespiratory fitness is more commonly expressed relative to body mass to account for differences in body size and to reflect the individuals´ ability to carry out weight-bearing tasks such as walking or running [6]. In this line, part of the association of VO_2max relative to body mass with the walk/run field tests reflects the influence of anthropometric variables [154]. Therefore, in the present meta-analysis is not surprising the fact that correlations of walk/run field tests with the VO_2max expressed relative to body mass is higher than those with the VO_2max expressed absolute or relative to fat-free mass. According to Meredith and Welk [9], the criterion-related validity of walk/run field tests with VO_2max relative to body mass should not be interpreted only in terms of cardiorespiratory fitness, but they also reflect the influence of differences on body fat. On the other hand, in the present meta-analysis statistically significant differences between the VO_2max relative to body mass and maximal aerobic speed were not found. The maximal aerobic speed defined as the lowest speed at which VO_2max occurs, besides the differences in body mass previously mentioned, it reflects other factors such as running economy. Although running economy is a metabolic factor that influences the running performance in a walk/run field test, it has shown not to increase the variance explained between the walk/run test score and the VO_2max relative to body mass [50].

Potentials and Limitations

The meta-analysis is a useful tool to assess scientific evidence, but an understanding of its potentials and limitations is needed for the most appropriate use of this method. An exhaustive review of the general potentials and limitations of meta-analyses, e.g. [11], as well as specifically in the meta-analysis of the criterion-related validity of cardiorespiratory fitness field tests has been published elsewhere [7]. Please see the strengths and limitations section of the above mentioned manuscript for more detailed information.

Regarding the potentials of the present meta-analysis, numerous measures to avoid, or at least to reduce, publication bias were followed. Then, several exploratory analyses were also conducted to identify and assess the impact of any potential publication bias. Another potential of the present meta-analysis is related to the statistical approach used. Since the Hunter-Schmidt's psychometric meta-analysis approach method [16] estimates the population correlation by correcting the observed correlations due to various artefacts such as sampling error and measurement error, it has been considered one of the best meta-analyses approaches [11].

As regards the limitations, the main ones were related to the small number of criterion-related validity coefficients found. Estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis. Additionally, because a partial hierarchical breakdown (instead of full) had to be used, misleading results due to confounding and interaction effects might be produced [16]. Therefore, the results of the present study should be considered with caution and firmer conclusions should await the accumulation of a larger number of studies [16].

Another potential limitation of the present meta-analysis could be related to the statistical metric used. Some authors have argued that the correlation coefficient is a measure of *relationship* rather than *agreement* and it might be also highly influenced by the range of individual measurements [155]. On the one hand, for instance, the agreement between the estimated VO_2max with an equation previously created with the same individuals and the actual VO_2max could be carried out. Unfortunately, since not many primary studies reported this information, the meta-analysis following the agreement approach simply could not be done. The performance score of the field tests (i.e. distance, time or speed) and the criterion measure (i.e. VO_2max) are expressed in two different units and, therefore, logically an agreement statistical approach could not be performed. In order to solve this potential methodological limitation, another kind of validity such as the cross-validity or criterion-referenced validity could be followed instead [156]. However, these approaches assess a different part of validity and they were not the scope of the present meta-analysis. For instance, the criterion-referenced validity simply examines the validity of a test for categorizing individuals in two or three categories. Although the criterion-referenced validity could be useful for some purpose such as screening if individuals are or not in a "health fitness zone", the criterion-related validity with continuous measurements is more appropriate for other common purposes such as analyzing the effects of an intervention program. Future research studies should examine the cross-validity and criterion-referenced validity of the walk/run field tests. Furthermore, the validity of other cardiorespiratory fitness field tests such as the walk or step tests should be also examined.

On the other hand, as regards the potential influence of the range of individuals´ measurement on criterion-related validity, the results of Spearman's rank order correlation between *r* values and standard deviation of the VO_2max did not show any statistically significant correlation for the walk/run field tests (5,000 m walk/run test: $r = -0.12$, $p = 0.651$; 3,000 m walk/run test: $r = 0.04$, $p = 0.907$; 1.5 miles walk/run test: $r = 0.19$, $p = 0.456$; 1 mile walk/run test: $r = 0.30$, $p = 0.106$; 12 min walk/run test: $r = 0.17$, $p = 0.449$; 9 min walk/run test: $r = -0.31$, $p = 0.355$), except for the 2 miles walk/run test ($r = 0.71$, $p = 0.009$). Therefore, in the present meta-analysis the empirical outcomes showed that the criterion-related validity values of the most distance- and time-based walk/run tests for estimating cardiorespiratory fitness were not biased by the variability of the sample measurements.

Another limitation of the present meta-analysis is related to the criterion measure used in the studies. Although only primary studies in which the criterion measure used the VO_2max relative to body mass during a laboratory-based incremental test to exhaustion were selected, researchers employed different equipment, ergometers and protocols, as well as criteria to determine VO_2max. Furthermore, it must be also highlighted the fact that in the present meta-analysis peak oxygen uptake (VO_2peak) has been used interchangeably with VO_2max. Although authors are aware that the VO_2peak simply refers to the highest value of oxygen uptake attained in a particular exercise test, due to the fact that the tests in the primary studies were maximal we can be reasonably sure that values were the highest value of oxygen uptake that is deemed attainable by individuals, i.e. the VO_2max [157].

Finally, coding some study features was problematic due to different reasons. Some study features simply could not be coded because the authors did not report them. Although authors were contacted by email and/or ResearchGate, many of them did not reply and the particular study feature had to be omitted. Also noting that many studies were published several years ago and, therefore, no contact email address and/or ResearchGate profile was found. On the other hand, because in the present meta-analysis the level of VO_2max was classified based on the average scores, readers have to be aware that several individuals with low VO_2max could be classified as high VO_2max and vice versa. This fact could affect the results of the present meta-analysis. Finally, although there could be other potentially moderating features such as physical activity levels or sport participation, coding for them was not possible because most studies did not report them.

Conclusions

The overall criterion-related validity of the distance- and time-based walk/run tests for estimating VO_2max ranged from low to moderate. Regarding the potential influence of the sex, age and fitness levels moderators, they seem not to affect the relationship between the walk/run tests scores and the measured VO_2max. According to the results of the present meta-analysis, the 1.5 mile and 12 min walk/run tests seem to be the best option of distance- and time-based tests, respectively. Meanwhile performing longer distances could be an unnecessary extra time and effort, shorter tests such as the 1 mile walk/run test showed poorer results of criterion-related validity for estimating cardiorespiratory fitness.

When an individual's VO_2max attained during a laboratory-based graded exercise test to exhaustion is not feasible like in sports clubs, physical education lessons or large scale research studies, scientists, practitioners and users could utilize the 1.5 mile or 12 min walk/run tests as useful alternatives to estimate cardiorespiratory fitness. Nevertheless, although among children the 1.5 mile and 12 min walk/run tests showed a similar criterion-related validity to the 20-m shuttle run test, among adults the 20-m shuttle run test should be used instead. As in the application of any physical fitness field test, evaluators must be aware that the performance score of the walk/run field tests is simply an estimation and not a direct measure of cardiorespiratory fitness.

Due to the relatively low number of *r* values found and the fact that the criterion-related validity of the walk/run tests within most categories is heterogeneous, we must be cautious with the results of the present meta-analysis. Therefore, when a greater number of studies is accumulated, a large sized meta-analysis with a full hierarchical analysis approach should be carried out. For this purpose future research studies should examine further the criterion-related validity of the walk/run tests. Furthermore, future research studies should examine the cross-validity and criterion-referenced validity of the walk/run tests. Finally, the validity of other cardiorespiratory fitness field tests such as the walk or step tests should be also examined.

Acknowledgements

We gratefully acknowledge all the authors of the original research studies for their contribution, without whom the present meta-analysis could not be done. Authors also thank the head of the library, Mrs. Ana M. Peregrín González, for her technical assistance and great help to retrieve manuscripts that were unavailable. Additionally, we thank the librarians Mrs. María Sagrario Avilés Rodríguez and Mr. Francisco Moldero

Espejo for their great help to digitalize some manuscripts in paper. Authors also thank Anna Szczesniak for the English revision.

References

1. Castillo Garzón MJ, Ortega Porcel FB, Ruiz Ruiz J. [Improvement of physical fitness as anti-aging intervention]. Med Clin (Barc). 2005;124:146-55. doi: 10.1157/13071011. Spanish.
2. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. J Appl Physiol. 1955;8:73-80.
3. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. JAMA. 2009;301:2024-35. doi: 10.1001/jama.2009.681
4. Blair SN. Physical inactivity: The biggest public health problem of the 21st century. Br J Sports Med. 2009;43:1-2.
5. Ruiz JR, Castro-Pinero J, Artero EG, Ortega FB, Sjostrom M, Suni J, et al. Predictive validity of health-related fitness in youth: A systematic review. Br J Sports Med. 2009;43:909-23. doi: 10.1136/bjsm.2008.056499
6. Pescatello LS, Arena R, Riebe D, Thompson PD. ACSM’s guidelines for exercise testing and prescription. 9th ed. Philadelphia: Wolters Kluwer/ Lippincott Williams & Wilkins; 2014.
7. Mayorga-Vega D, Aguilar-Soto P, Viciana J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. J Sports Sci Med. 2015;14:536-47.

8. Castro-Piñero J, Artero EG, España-Romero V, Ortega FB, Sjöström M, Ruiz JR. Criterion-related validity of field-based fitness tests in youth: A systematic review. Br J Sports Med. 2010;44:934-43. doi: 10.1136/bjsm.2009.058321

9. Meredith MD, Welk GJ, editors. FITNESSGRAM® & ACTIVITYGRAM® test administration manual. 4th ed. Champaign: Human Kinetics; 2010.

10. Ruiz JR, Ortega FB, Castro-Piñero J. Validity and reliability of the 1/4 mile run-walk test in physically active children and adolescents. Nutr Hosp. 2015;31:875-82. doi: 10.3305/nh.2015.31.2.7789

11. Cooper H, Hedges LV, Valentine JC. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Sage; 2009.

12. Cooper KH. A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. J Am Med Assoc. 1968;203:201-4.

13. O'Gorman D, Hunter A, McDonnacha C, Kirwan J. Validity of field tests for evaluating endurance capacity in competitive and international-level sports participants. J Strength Cond Res. 2000;14:62-7.

14. Casajus JA, Castagna C. Aerobic fitness and field test performance in elite Spanish soccer referees of different ages. J Sci Med Sport. 2007;10:382-9. doi: 10.1016/j.jsams.2006.08.004

15. Lipsey MW, Wilson DB. Practical meta-analysis. Newbury Park: Sage; 2001.

16. Schmidt FL, Hunter JE. Methods of meta-analysis: Correcting error and bias in research findings. 3nd ed. California: Sage; 2015.

17. Mayorga-Vega D, Merino-Marban R, Viciana J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. J Sports Sci Med. 2014;13:1-14.

18. Mayorga-Vega D, Viciana J, Cocca A, Merino-Marban R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis. J Human Sport Exerc. 2014;9:188-200.

19. Safrit MJ, Hooper LM, Ehlert SA, Costa MG, Patterson P. The validity generalization of distance run tests. Can J Sport Sci. 1988;13:188-96.

20. Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: Larger studies may be more reliable. Control Clin Trials. 1997;18:568-79.

21. Light RJ, Pillemer DB. Summing up: The science of reviewing research. Cambridge: Harvard University Press; 1984.

22. Begg CB, Mazumdar M. Operating characteristics of a rank order correlation for publication bias. Biometrics. 1994;50:1088-101.

23. Orwin RG. A fail-safe n for effect size in meta-analysis. J Educ Stat. 1983;8:157-9.

24. Cohen JA. Power primer. Psychol Bull. 1992;112:155-9.

25. Almarwaey OA, Jones AM, Tolfrey K. Physiological correlates with endurance performance in trained adolescents. Med Sci Sports Exerc. 2003;35:480-7. doi: 10.1249/01.MSS.0000053723.16531.D0

26. Anderson GS. The 1600-m run and multistage 20-m shuttle run as predictive tests of aerobic capacity in children. Pediatr Exerc Sci. 1992;4:312-8.

27. Arabas JL, Anderson MMEP, Arabas JR, Arabas CD, Mayhew JL.. Estimation of VO_2max from 9-minute run performance. Am J Health Educ. 1996;29:11.

28. Baldwin DL. Prediction of VO_2 values from 9-minute run distances in young males, 9-14 years [dissertation]. La Crosse: University of Wisconsin-La Crosse; 1983.

29. Bandyopadhyay A. Validity of Cooper's 12-min run test for estimation of maximum oxygen uptake in female university students. Indian J Physiol Pharmacol. 2014;58:184-6.

30. Bandyopadhyay A. Validity of Cooper's 12-minute run test for estimation of maximum oxygen uptake in male university students. Biol Sport. 2015;32:59-63. doi: 10.5604/20831862.1127283

31. Barbineau C, Léger L. Physiological response of 5/1 intermittent aerobic exercise and its relationship to 5 km endurance performance. Int J Sports Med. 1997;18:13-9. doi: 10.1055/s-2007-972588

32. Bergmann GG, de Araújo Bergmann ML, Mattos de Castro AA, Del´Corona Lorenzi T, dos Santos Pinheiro E, Moreira RB, et al. Use of the 6-minute walk/run test to predict peak oxygen uptake in adolescents. Rev Bras Ativid Fís Saúde. 2014;19:64-73. doi: 10.12820/rbafs.v.19n1p64

33. Bergmann GG, de Araújo Bergmann ML, Mattos de Castro AA, Del'Corona Lorenzi T, dos Santos Pinheiro E, Moreira RB, et al. Prediction of peak oxygen uptake in adolescents from 9 minutes run/walk test. Gazz Med Ital. 2015;174:15-22.

34. Borghols EAM, Dresen MHW, Poulus AJ. [Cooper test: Useful or not?]. Geneeskd Sport. 1981;14:74-7. Dutch.

35. Brandon LJ, Boileau RA. Influence of metabolic, mechanical and physique variables on middle distance running. J Sports Med Phys Fitness. 1992;32:1-9.

36. Buono MJ, Roby JJ, Micale FG, Sallis JF, Shepard WE. Validity and reliability of predicting maximum oxygen uptake via field tests in children and adolescents. Pediatr Exerc Sci. 1991;3:250-5.

37. Burger SC, Bertram SR, Stewart RI. Assessment of the 2,4 km run as a predictor of aerobic capacity. S Afr Med J. 1990;78:327-9.

38. Burke EJ. Validity of selected laboratory and field tests of physical working capacity. Res Q. 1976;47:95-104.

39. Burns RD, Hannon JC, Brusseau TA, Eisenman PA, Saint-Maurice PF, Welk G, et al. Cross-validation of aerobic capacity prediction models in adolescents. Pediatr Exerc Sci. 2015;27:404-11. doi: 10.1123/pes.2014-0175

40. Burns RD, Hannon JC, Brusseau TA, Eisenman PA, Shultz BB, Saint-Maurice PF, et al. Development of an aerobic capacity prediction model from one-mile run/walk performance in adolescents aged 13-16 years. J Sports Sci. Forthcoming 2015. doi: 10.1080/02640414.2015.1031163

41. Calders P, Deforche B, Verschelde S, Bouckaert J, Chevalier F, Bassle E, et al. Predictors of 6-minute walk test and 12-minute walk/run test in obese children and adolescents. Eur J Pediatr. 2008;167:563-8. doi: 10.1007/s00431-007-0553-5

42. Castro-Piñero J, Mora J, González-Montesinos JL, Sjöström M, Ruiz JR. Criterion-related validity of the one-mile run/walk test in children aged 8-17 years. J Sports Sci. 2009;27:405-13. doi: 10.1080/02640410802603889

43. Castro-Piñero J, Ortega FB, Mora J, Sjöström M, Ruiz JR. Criterion related validity of 1/2 mile run-walk test for estimating VO_2peak in children aged 6-17 years. Int J Sports Med. 2009;30:366-71. doi: 10.1055/s-0028-1105934

44. Cisar CJ, Thorland WG, Johnson GO, Housh TJ. The effect of endurance training on metabolic responses and the prediction of distance running performance. J Sports Med Phys Fitness. 1986;26:234-40.

45. Chiou LA. A study of correlationship among maximal oxygen uptake, anaerobic threshold vs 12 minutes running and 400m running perfomance. Asian J Phys Educ. 1991;14:32-3.

46. Conley DS, Cureton KJ, Dengel DR, Weyand PG. Validation of the 12-min swim as a field test of peak aerobic power in young men. Med Sci Sports Exerc. 1991;23:766-73.

47. Coolbaugh CL, Anderson IB, Wilson MD, Hawkins DA, Amsterdam EA. Evaluation of an exercise field test using heart rate monitors to assess cardiorespiratory fitness and heart rate recovery in an asymptomatic population. PLoS One. 2014;9:e97704. doi: 10.1371/journal.pone.0097704

48. Cunningham LN, Cama G, Cilia G, Bazzano C. Relationship of VO_2max with the 1-mile run and 20 meter shuttle test VO_2 with youth aged 11 to 14 years. Med Sci Sports Exerc. 1994;26:S209.

49. Cureton KJ, Boileau RA, Lohman TG, Misner JE. Determinants of distance running performance in children: Analysis of a path model. Res Q. 1977 ;48:270-9.

50. Cureton KJ, Sloniger MA, Black DM, McCormack WP, Rowe DA. Metabolic determinants of the age-related improvement in one-mile run/walk performance in youth. Med Sci Sports Exerc. 1997;29:259-67.

51. Cureton KJ, Sloniger MA, O'Bannon JP, Black DM, McCormack WP. A generalized equation for prediction of VO_2peak from 1-mile run/walk performance. Med Sci Sports Exerc. 1995;27:445-51.

52. Da Silva SG, Moyna NM, Robertson RJ, Goss FL, Metz KF. Prediction of 3000m endurance performance in male and female distance runners using velocity at VO_2max. Med Sci Sports Exerc. 1992;24:S102.

53. Dal Pupo J, Arins FB, Guglielmo LGA, Da Silva RCR, Moro ARP, Dos Santos SG. Physiological and neuromuscular indices associated with sprint running performance. Res Sports Med. 2013;21:124-35. doi: 10.1080/15438627.2012.757225

54. Damitz SR, Ebbeling CE, Ward A, Freedson P, Rippe JM. Validity of the one mile run/walk test in children ages 6 to 13 years. Med Sci Sports Exerc. 1994;26:S209.

55. De Almeida JA, Campbell CSG, Pardono E, Sotero RC, Magalhães G, Simões HG. [Predictive equations validity in estimating the VO_2max of young Brazilians from performance in a 1600 m run]. Rev Bras Med Esporte. 2010;16:57-60. Portuguese.

56. Díaz FJ, Montaño JG, Melchor MT, Guerrero JH, Tovar JA. [Validation and confiability of aerobic test of 1,000 meters]. Rev Invest Clin. 2000;52:44-51. Spanish.

57. Dorociak JJ. Validity of running tests of 4, 8, and 12 minutes duration in estimating aerobic power for college women of different fitness levels [dissertation]. South Carolina: University of South Carolina; 1981.

58. Dorociak JJ, Nelson JK. The 1 mile and 2 mile runs as measures of cardiovascular fitness in college women. J Sports Med Phys Fitness. 1983;23:322-5.

59. Drinkard B, McDuffie J, McCann S, Uwaifo GI, Nicholson J, Yanovski JA. Relationships between walk/run performance and cardiorespiratory fitness in adolescents who are overweight. Phys Therapy. 2001;81:1889-96.

60. Epperson CE, Buono MJ, Kolkhorst FW, Reynolds KK, Nanista JA, Sheffield RD. Correlation of lactate threshold, VO_2max, and running economy with 4.8 kilometer running performance. Med Sci Sports Exerc. 1999;31:S104.

61. Fay L, Londeree BR, LaFontaine TP, Volek MR. Physiological parameters related to distance running performance in female athletes. Med Sci Sports Exerc. 1989;21:319-24.

62. Ferfila N, Milić R, Škof B. [Influence of the energy profile of children on their results in the 600-metre run]. Sport: Revija Za Teoreticna in Prakticna Vprasanja Sporta. 2014;62:201-6. Slovenian.

63. Fontana KE. [Comparison of VO_2max in direct and indirect measurements on treadmill and field test]. Rev Bras Ciênc Esporte. 1983;4:78-90. Portuguese.

64. Foster C, Costill DL, Daniels JT, Fink WJ. Skeletal muscle enzyme activity, fiber composition and VO2 max in relation to distance running performance. Eur J Appl Physiol Occup Physiol. 1978;39:73-80.

65. George JD, Larsen GE, Alexander JL, Fellingham GW, Aldana SG. Combining walking, jogging, and running into a single VO_2max prediction test. Med Sci Sports Exerc. 2001;33:S45.

66. George JD, Vehrs PR, Allsen PE, Fellingham GW, Fisher AG. VO_2max estimation from a submaximal 1-mile track jog for fit college-age individuals. Med Sci Sports Exerc. 1993;25:401-6.

67. Getchell LH, Kirkendall D, Robbins G. Prediction of maximal oxygen uptake in young adult women joggers. Res Q. 1977;48:61-7.

68. Ghosh AK, Ahuja A, Khanna GL. Distance run as a predictor of aerobic endurance (VO_2max) of sportsmen. Indian J Med Res. 1987;85:680-4.

69. Grant S, Craig I, Wilson J, Aitchison T. The relationship between 3 km running performance and selected physiological variables. J Sports Sci. 1997;15:403-10. doi: 10.1080/026404197367191

70. Grant JA, Joseph AN, Campagna PD. The prediction of VO_2max: A comparison of 7 indirect tests of aerobic power. J Strength Cond Res. 1999;13:346-52.

71. Guo-jun W, Han W, Ya-qiong W. [Running economy versus maximal oxygen uptake to evaluate the endurance level of ordinary people]. Chin J Tissue Engineering Res. 2013;17:1265-72. doi: 10.3969/j.issn.2095-4344.2013.07.022. Chinese.

72. Gutin B, Fogle RK, Stewart K. Relationship among submaximal heart rate, aerobic power, and running performance in children. Res Q. 1976;47:536-9.

73. Gutin B, Torrey K, Welles R, Vytvytsky M. Physiological parameters related to running performance in college trackmen. J Hum Ergol (Tokyo). 1975;4:27-34.

74. Haines DA, Wilby K. Relationship between lung function and physical fitness in 9 to 15 year old Australian children. Aust J Sci Med Sport. 1993;25:35-9.

75. Hamlin MJ, Fraser M, Lizamore CA, Draper N, Shearman JP, Kimber NE. Measurement of cardiorespiratory fitness in children from two commonly used field tests after accounting for body fatness and maturity. J Hum Kinet. 2014;40:83-92. doi: 10.2478/hukin-2014-0010

76. Hergenroeder AC, Schoene RB. Predicting maximum oxygen uptake in adolescents. Am J Dis Child. 1989;143:673-7.

77. Howald H, Ehrsam R, Rüegger B, Meienhofer R, Mohler H, Oertli M, et al. [Conception and evaluation of a simple fitness test]. Schweiz Z Sportmed. 1975;23:57-92. German.

78. Ingham SA, Whyte GP, Pedlar C, Bailey DM, Dunman N, Nevill AM. Determinants of 800-m and 1500-m running performance using allometric models. Med Sci Sports Exerc. 2008;40:345-50. doi: 10.1249/mss.0b013e31815a83dc

79. Ishiko T. Aerobic capacity and external criteria of performance. CMAJ. 1967;96:746-9.

80. Jackson AS, Coleman AE. Validation of distance run tests for elementary school children. Res Q. 1976;47:86-94.

81. Jackson A, der Weduwe K, Schick R, Sánchez R. An analysis of the validity of the three-mile run as a field test of aerobic capacity in college males. Res Q Exerc Sport. 1990;61:233-7.

82. Jessup GT, Tolson H, Terry JW. Prediction of maximal oxygen intake from Astrand-Rhyming test, 12-minute, and anthropometric variables using stepwise multiple regression. Am J Phys Med. 1974;53:200-7.

83. Katch FI, McArdle WD, Czula R, Pechar GS. Maximal oxygen intake, endurance running performance, and body composition in college women. Res Q. 1973;44:301-12.

84. Kirk SM, Weiglein L, Herrick J, Kirk EP. The one-mile walk and 1.5 mile run are valid assessments of cardiovascular fitness in air force servicemen. Med Sci Sports Exerc. 2014;46:S116.

85. Kitagawa K, Yamamoto K, Miyashita M. Maximal oxygen uptake, body composition and running performance in Japanese young adults of both sexes. In: Landry F, Orban WAR, editors: Proceedings of the International Congress of Physical Activity Sciences; 1976; Quebec, Canada: Symposia Specialist; 1978. p. 553-61.

86. Krahenbuhl GS, Pangrazi RP. Characteristics associated with running performance in young boys. Med Sci Sports Exerc. 1983;15:486-90.

87. Krahenbuhl GS, Pangrazi RP, Burkett LN, Schneider MJ, Petersen G. Field estimation of VO_2max in children eight years of age. Med Sci Sports. 1977;9:37-40.

88. Kumagai S, Tanaka K, Matsuura Y, Matsuzaka A, Hirakoba K, Asano K. Relationships of the anaerobic threshold with the 5 km, 10 km, and 10 mile races. Eur J Appl Physiol Occup Physiol. 1982;49:13-23.

89. Larsen GE, George JD, Alexander JL, Fellingham GW, Aldana SG, Parcell AC. Prediction of maximum oxygen consumption from walking, jogging, or running. Res Q Exerc Sport. 2002;73:66-72. doi: 10.1080/02701367.2002.10608993

90. Lawrenz W, Stemper T. [Comparison of the 6-minute-jog-walk and maximal oxygen uptake in 8-10-year old school children]. Dtsch Z Sportmed. 2012;63:102-5. doi: 10.5960/dzsm.2012.012. German.

91. Li LW, Shang YJ, Chorng LJ. Relationship between twelve minute walk-run and maximal oxygen consumption in university male students. Asian J Phys Educ. 1985;14:46-7.

92. Mahon AD, Del Corral P, Howe CA, Duncan GE, Ray ML. Physiological correlates of 3-kilometer running performance in male children. Int J Sports Med. 1996;17:580-4. doi: 10.1055/s-2007-972898

93. Marsh HW, Ridge BR. The construct validity and generalisability of VO_2max for boys and girls aged 9-15. Aust J Sci Med Sport. 1993;25:73-9.

94. Massicotte DR, Gauthier R, Markon P. Prediction of VO_2max from the running performance in children aged 10-17 years. J Sports Med Phys Fitness. 1985 ;25:10-7.

95. Mayhew JL, Andrew J. Assessment of running performance in college males from aerobic capacity percentage utilization coefficients. J Sports Med Phys Fitness. 1975;15:342-6.
96. McCormack WP, Cureton KJ, Bullock TA, Weyand PG. Metabolic determinants of 1-mile run/walk performance in children. Med Sci Sports Exerc. 1991;23:611-7.
97. McCreight GA. Field estimation of cardiorespiratory fitness in young females, eight to eleven years of age [dissertation]. Vancouver: The University of British Columbia; 1982.
98. Mello RP, Murphy MM, Vogel JA. Relationship between the army two mile run test and maximal oxygen uptake. Fort Detrick (US): Army Research Institute of Environmental Medicine; 1984 Dec. Report No.: T3/85.
99. Myles WS, Brown TE, Pope JI. A reassessment of a running test as a measure of cardiorespiratory fitness. Ergonomics. 1980;23:543-7.
100. Nevill AM, Ramsbottom R, Nevill ME, Newport S, Williams C. The relative contributions of anaerobic and aerobic energy supply during track 100-, 400-, and 800-m performance. J Sports Med Phys Fitness. 2008;48:138-42.
101. Ong TC, Sothy SP. Exercise and cardiorespiratory fitness. Ergonomics. 1986;29:273-80.
102. Paludo AC, Batista MB, Gobbo LA, Ronque ERV, Petroski EL, Serassuelo H. [Development of equations to estimate the VO_2peak by the 9-minute test]. Rev Bras Med Esporte. 2014;20:176-80. doi: 10.1590/1517-86922014200301564. Portuguese.
103. Paludo AC, Batista MB, Serassuelo H, Cyrino ES, Ronque ERV. Estimation of cardiorespiratory fitness in adolescents with the 9-minute run/walk test. Rev

Bras Cineantrop Desemp Hum. 2012;14:401-8. doi: 10.5007/1980-0037.2012v14n4p401

104. Penry JT. Validity and reliability analysis of Cooper's 12-minute run and the multistage shuttle run in healthy adults [dissertation]. Oregon: Oregon State University;2008.

105. Plowman SA, Liu NYS. Norm-referenced and criterion-referenced validity of the one-mile run and PACER in college age individuals. Meas Phys Educ Exerc Sci. 1999;3:63-84. doi: 10.1207/s15327841mpee0302_1

106. Pomerants T. Indirect measurement of maximal oxygen consumption in 14-16-year old adolescent girls [dissertation]. Tartu: University of Tartu; 2004.

107. Quinart S, Mougin F, Simon-Rigaud ML, Nicolet-Guénat M, Nègre V, Regnard J. Evaluation of cardiorespiratory fitness using three field tests in obese adolescents: Validity, sensitivity and prediction of peak VO_2. J Sci Med Sport. 2014;17:521-5. doi: 10.1016/j.jsams.2013.07.010

108. Ramsbottom R, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. Br J Sports Med. 1988;22:141-4. doi: 10.1136/bjsm.22.4.141

109. Ramsbottom R, Nute MGL, Williams C. Determinants of five kilometre running performance in active men and women. Br J Sports Med. 1987; 21:9-13. doi: 10.1136/bjsm.21.2.9

110. Ramsbottom R, Williams C, Boobis L, Freeman W. Aerobic fitness and running performance of male and female recreational runners. J Sports Sci. 1989;7:9-20.

111. Ramsbottom R, Williams C, Kerwin DG, Nute MLG. Physiological and metabolic responses of men and women to a 5-km treadmill time trial. J Sports Sci. 1992;10:119-29.

112. Rasch PJ. Maximal oxygen intake as a predictor of performance in running events. J Sports Med Phys Fitness. 1974;14:32-9.

113. Redkva PE, Zagatto AM, Gomes EB, Kalva-Filho CA, Loures JP, Kaminagakura EI, et al. Prediction of aerobic performance in distance from 1200 to 2800 m for laboratory testing with military runners. J Exerc Physiol. 2012;15:107-14.

114. Rowland T, Kline G, Goff D, Martel L, Ferrone L. One-mile run performance and cardiovascular fitness in children. Arch Pediatr Adolesc Med. 1999;153:845-9.

115. Scott BK, Houmard JA. Peak running velocity is highly related to distance running performance. Int J Sports Med. 1994;15:504-7.

116. Slattery KM, Wallace LK, Murphy AJ, Coutts A. Physiological determinants of three-kilometer running performance in experienced triathletes. J Strength Cond Res. 2006;20:47-52.

117. Sloniger MA, Cureton KJ, O'Bannon PJ. One-mile run-walk performance in young men and women: Role of anaerobic metabolism. Can J Appl Physiol. 1997;22:337-50.

118. Soong PX, Leng YW, Hu QC. Validation of 1600-m run as a predictor of VO_2max in Singapore children aged 10-12 years. Med Sci Sports Exerc. 2005;37:S17.

119. Sparling PB, Cureton KJ. Biological determinants of the sex difference in 12-min run performance. Med Sci Sports Exerc. 1983;15:218-23.

120. Sporiš G. Validity of 2-mile run test for determination of VO_2max among soldiers. J Sport Hum Perform. 2013;1:15-22. doi: 10.12922/jshp.v1i1.4

121. Sporiš G, Harasin D, Baić M, Krističević T, Krakan I, Milanović Z, et al. The effects of basic fitness parameters on the implementation of specific military activities. Coll Antropol. 2014;38:165-71.

122. Stratton E, O'Brien BJ, Harvey J, Blitvich J, McNicol AJ, Janissen D, et al. Treadmill velocity best predicts 5000-m run performance. Int J Sports Med. 2009;30:40-5. doi: 10.1055/s-2008-1038761

123. Sucec AA, Black DG, Mittleman KD. The interrelationship of VO_2max, anaerobic threshold and 15 minute run performance. Int J Sports Med. 1982;3:66.

124. Sucec AA, Burks J, Buffington S, Trone D. Gender differences in prediction accuracy for lactate threshold and VO_2max for middle distance races. Med Sci Sports Exerc. 1999;31:S379.

125. Suh SH, Kim KB. [Validity analysis and estimation for aerobic fitness of 3000m distance run in military fitness test]. Korean J Meas Eval Phys Educ Sport Sci. 2010;12:25-33. Korean.

126. Swisher AK, Goldfarb AH. Use of the six-minute walk/run test to predict peak oxygen consumption in older adults. Cardiopulm Phys Ther J. 1998;9:3-5.

127. Tanaka K, Matsuura Y, Moritani T. A correlational analysis of maximal oxygen uptake and anaerobic threshold as compared with middle and long distance performances. Jpn J Phys Fit Sports Med. 1981;30:94-102. doi: 10.7600/jspfsm1949.30.94

128. Tanaka K, Nakadomo F, Maeda K. Effects of jogging/running training on cardiorespiratory fitness, serum lipids, and body composition in healthy males. Ann Physiol Anthropol. 1988;7:31-8.

129. Tanaka K, Takeshima N, Kato T, Niihata S, Ueda K. Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. Eur J Appl Physiol Occup Physiol. 1990;59:443-9.

130. Turley KR, Wilmore JH, Simons-Morton B, Williston JM, Epping JR, Dahlstrom G. The reliability and validity of the 9-minute run in third-grade children. Pediatr Exerc Sci. 1994;6:178-87.

131. Unnithan VB, Timmons JA, Paton JY, Rowland TW. Physiologic correlates to running performance in pre-pubertal distance runners. Int J Sports Med. 1995;16:528-33. doi: 10.1055/s-2007-973049

132. Van Mechelen W, Hlobil H, Kemper HCG. Validation of two running tests as estimates of maximal aerobic power in children. Eur J Appl Physiol Occup Physiol. 1986;55:503-6.

133. Vehrs PR, Quail JW, Jackson AS. VO_2max estimation from self administration of the 1-mile track jog test. Med Sci Sports Exerc. 1996;28:S182.

134. Vodak PA, Wilmore JH. Validity of the 6-minute jog-walk and the 600-yard run-walk in estimating endurance capacity in boys, 9-12 years of age. Res Q. 1975;46:230-4.

135. Von Haaren B, Härtel S, Seidel I, Schlenker L, Bös K. [Validity of a 6-min endurance run and a 20-m shuttle run in 9- to 11-year old children]. Dtsch Z Sportmed. 2011;62:351-5. German.

136. Walker LA, Sharp MA, Knapik JJ, Marín RE, Mello RP. Army physical fitness test 2-mile run correlates with peak oxygen uptake in infantry soldiers. Med Sci Sports Exerc. 2009;41:S101-S102. doi: 10.1249/01.MSS.0000354868.79967.47

137. Weiglein L, Herrick J, Kirk S, Kirk EP. The 1-mile walk test is a valid predictor of VO_2max and is a reliable alternative fitness test to the 1.5-mile run in U.S Air Force males. Mil Med. 2011;176:669-73.

138. Weltman A, Seip R, Bogardus AJ, Snead D, Dowling E, Levine S, et al. Prediction of lactate threshold (LT) and fixed blood lactate concentrations (FBLC) from 3200-m running performance in women. Int J Sports Med. 1990;11:373-8.

139. Weltman J, Seip R, Levine S, Snead D, Rogol A, Weltman A. Prediction of lactate threshold and fixed blood lactate concentrations from 3200-m time trial running performance in untrained females. Int J Sports Med. 1989;10:207-11.

140. Weltman A, Snead D, Seip R, Schurrer R, Levine S, Rutt R, et al. Prediction of lactate threshold and fixed blood lactate concentrations from 3200-m running performance in male runners. Int J Sports Med. 1987;8:401-6.

141. Weyand PG, Cureton KJ, Conley DS, Sloniger MA, Liu YL. Peak oxygen deficit predicts sprint and middle-distance track performance. Med Sci Sports Exerc. 1994;26:1174-80.

142. Wiley JF, Shaver LG. Prediction of maximum oxygen intake from running performances of untrained young men. Res Q. 1972;43:89-93.

143. Yoshida T, Ishiko T. Physiological studies on cardiorespiratory response to exercise and validity of endurance tests in ten-year-old boys. In: Landry F, Orban WAR, editors: Proceedings of the International Congress of Physical Activity Sciences; 1976; Quebec, Canada: Symposia Specialist; 1978. p. 541-5.

144. Burns RD. Development and cross-validation of aerobic capacity prediction models in adolescent youth [dissertation]. Utah: University of Utah; 2014.

145. Castro-Piñero J, Mora J, González-Montesinos JL, Sjöström M, Ruiz JR. Criterion-related validity of 1-mile run/walk test in children aged 8-17 years. Med Sci Sports Exerc. 2009;41:S121. doi: 10.1249/01.mss.0000353656.17328.b5

146. Hamlin MJ, Fraser M, Lizamore CA, Fryer S, Draper N, Shearman JP, Kimber NE. Criterion-related validity of the 20-m shuttle and the 550-m distance run in 8-13 year-old children. Proceedings of the British Association for Sport and Exercise Science 2013 Conference; 2013 Sep 3-5; Preston; 2013

147. Terrin N, Schmid CP, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58:894-901.

148. Cadenas-Sánchez C, Alcántara-Moral F, Sánchez-Delgado G, Mora-González J, Martínez-Téllez B, Herrador-Colmenero M, et al. [Assessment of cardiorespiratory fitness in preschool children: Adaptation of the 20 metres shuttle run test]. Nutr Hosp. 2014;30:1333-43. doi: 0.3305/nh.2014.30.6.7859. Spanish.

149. Midgley AW, Bentley DJ, Luttikholt H, McNaughton LR, Millet GP. Challenging a dogma of exercise physiology. Does an incremental exercise test for valid VO_2max determination really need to last between 8 and 12 minutes? Sports Med. 2008;38:441-7.

150. Castro-Piñero J, Ortega FB, Keating XD, González-Montesinos JL, Sjöstrom M, Ruiz JR. Percentile values for aerobic performance running/walking field tests in children aged 6 to 17 years; influence of weight status. Nutr Hosp. 2011; 26:572-8. doi: 10.3305/nh.2011.26.3.4597

151. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 meter shuttle run test for aerobic fitness. J Sports Sci. 1988;6:93-101.

152. Council of Europe Committee for the Development of Sport. Eurofit Handbook for the EUROFIT tests of physical fitness. Rome: Edigraf Editoriale Grafica; 1988.

153. Riddoch CJ. The Northern Ireland health and fitness survey-1989: The fitness, physical activity, attitudes and lifestyles of Northern Ireland post-primary schoolchildren. The Queen's University of Belfast, Belfast; 1990.

154. Fernhall B, Tymeson GT. Validation of cardiovascular fitness field tests for adults with mental retardation. Adapt Phys Activ Q. 1988;5:49-59.

155. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26:217-38.

156. Baumgartner TA, Jackson AS, Mahar MT, Rowe DA. Measurement for evaluation in kinesiology. 9th ed. Burlington: Jones & Bartlett Learning; 2015.

157. Rowland TW. Does peak VO_2 reflect VO_2max in children?: Evidence from supramaximal testing. Med Sci Sports Exerc. 1993;25:689-93.

Supporting Information

S1 Appendix. Syntaxes used in the present study for the search with the electronic bibliographic databases.

S1 Table. Summary of the included studies examining the criterion-related validity of walk/run field tests for estimating cardiorespiratory fitness

S2 Table. Checklist.

Figure 1

Click here to download Figure Fig 1.tif

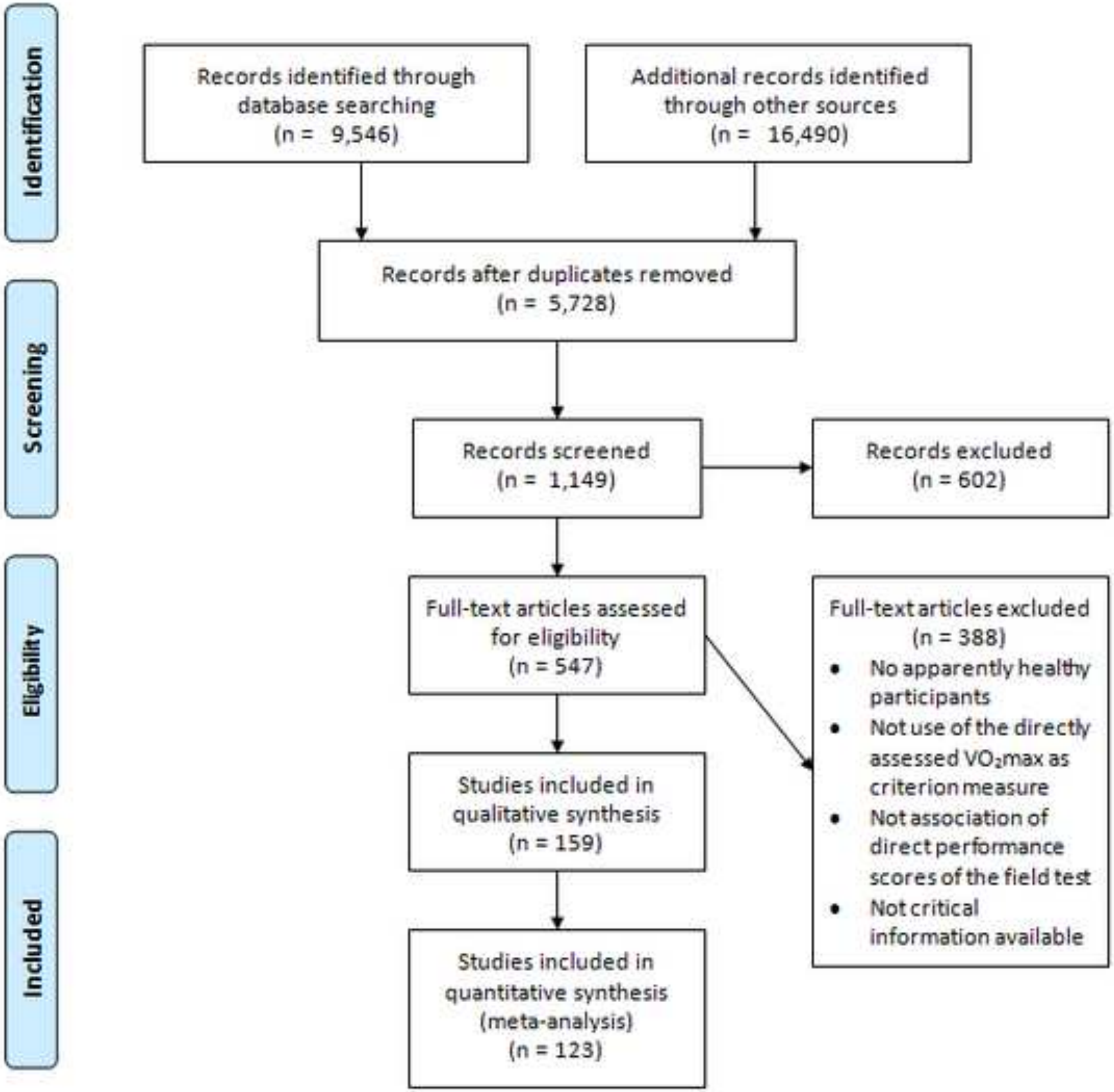


Figure 2

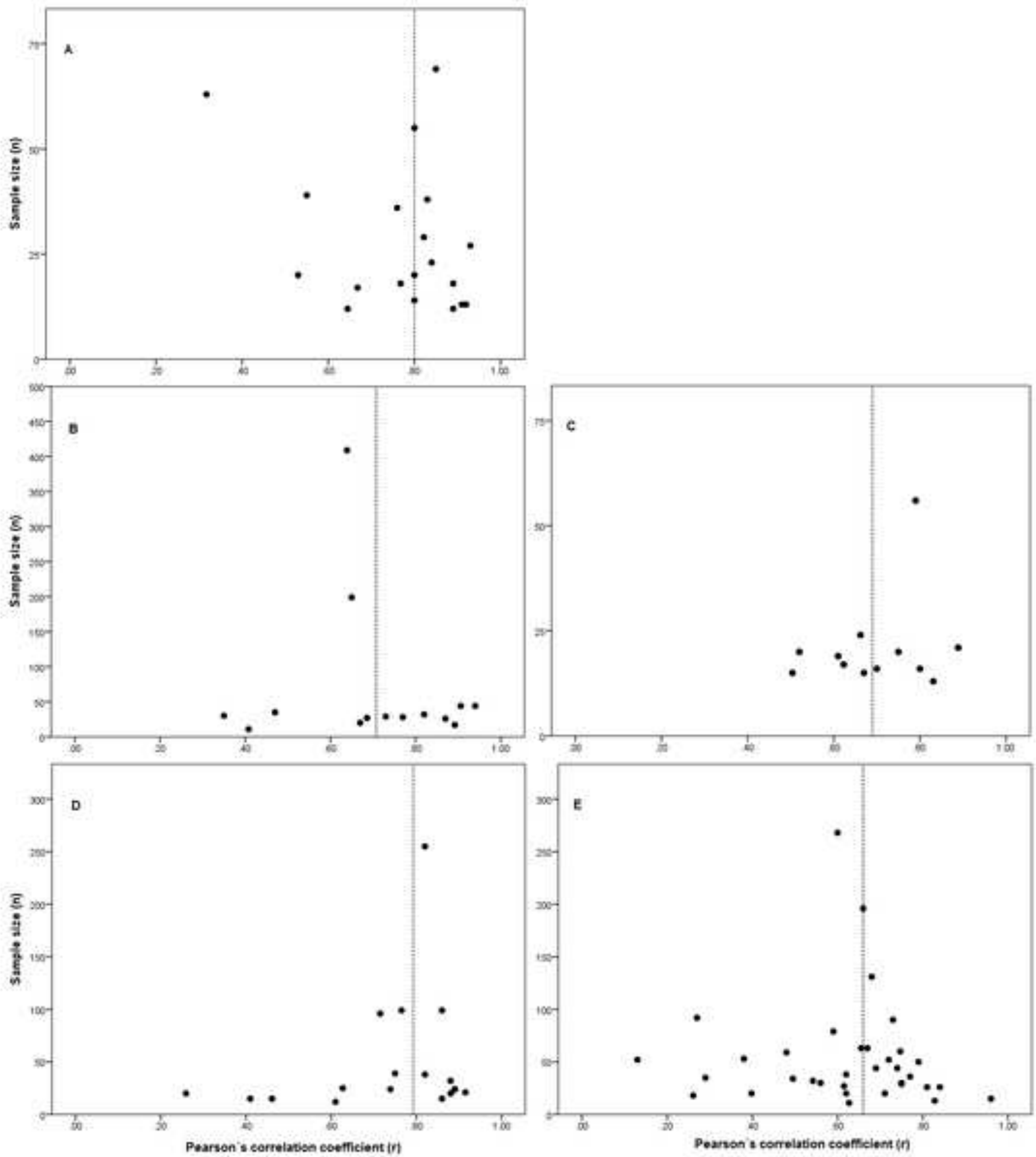
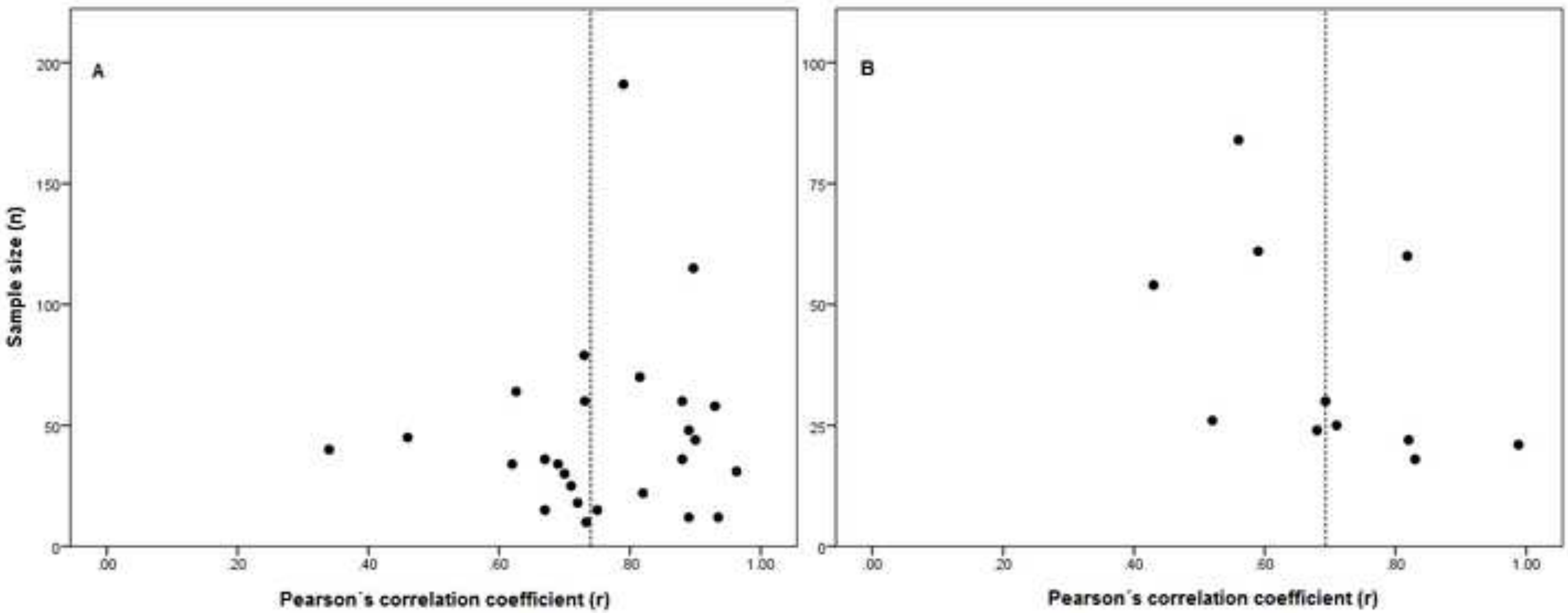

A
B
C
D
E
Sample size (n)
Pearson´s correlation coefficient (r)
.00
.20
.40
.60
.80
1.00

Figure 3


A
B
Sample size (n)
200
150
100
50
0
75
25
.00
.20
.40
.60
.80
1.00
Pearson´s correlation coefficient (r)

Click here to access/download
Supporting Information - Compressed/ZIP File Archive

Click here to access/download

Supporting Information - Compressed/ZIP File Archive

Click here to access/download
Supporting Information - Compressed/ZIP File Archive

Paper VII

CRITERION-RELATED VALIDITY OF SIT-AND-REACH TESTS FOR ESTIMATING HAMSTRING AND LUMBAR EXTENSIBILITY: A META-ANALYSIS

Mayorga-Vega, D., Merino-Marban, R., & Viciana, J.

Journal of Sports Science and Medicine

2014, *13*(1), 1-14.

©Journal of Sports Science and Medicine (2014) **13**, 01-14
http://www.jssm.org

Review article

Criterion-Related Validity of Sit-And-Reach Tests for Estimating Hamstring and Lumbar Extensibility: A Meta-Analysis

Daniel Mayorga-Vega [1]✉, Rafael Merino-Marban [2] and Jesús Viciana [1]
[1] Department of Physical Education and Sport, University of Granada, Spain; [2] Department of Didactics of Musical, Plastic and Corporal Expression, University of Malaga, Spain

Abstract
The main purpose of the present meta-analysis was to examine the scientific literature on the criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility. For this purpose relevant studies were searched from seven electronic databases dated up through December 2012. Primary outcomes of criterion-related validity were Pearson´s zero-order correlation coefficients (r) between sit-and-reach tests and hamstrings and/or lumbar extensibility criterion measures. Then, from the included studies, the Hunter-Schmidt´s psychometric meta-analysis approach was conducted to estimate population criterion-related validity of sit-and-reach tests. Firstly, the corrected correlation mean (r_p), unaffected by statistical artefacts (i.e., sampling error and measurement error), was calculated separately for each sit-and-reach test. Subsequently, the three potential moderator variables (sex of participants, age of participants, and level of hamstring extensibility) were examined by a partially hierarchical analysis. Of the 34 studies included in the present meta-analysis, 99 correlations values across eight sit-and-reach tests and 51 across seven sit-and-reach tests were retrieved for hamstring and lumbar extensibility, respectively. The overall results showed that all sit-and-reach tests had a moderate mean criterion-related validity for estimating hamstring extensibility (r_p = 0.46-0.67), but they had a low mean for estimating lumbar extensibility (r_p = 0.16-0.35). Generally, females, adults and participants with high levels of hamstring extensibility tended to have greater mean values of criterion-related validity for estimating hamstring extensibility. When the use of angular tests is limited such as in a school setting or in large scale studies, scientists and practitioners could use the sit-and-reach tests as a useful alternative for hamstring extensibility estimation, but not for estimating lumbar extensibility.

Key words: Concurrent validity, range of motion, flexibility, field test, lineal test, systematic review.

Introduction

Lack of hamstring muscles extensibility conditions a decrease of pelvic mobility (Kendall et al., 2005). This invariably leads to biomechanical changes in the pressure distribution of the spine and consequent spinal disorders (da Silva Días and Gómez-Conesa, 2008). Therefore, poor hamstring extensibility has been associated with thoracic hyperkyphosis (Fisk et al., 1984), spondylolysis (Standaert and Herring, 2000), disc herniation (Harvey and Tanner, 1991), changes in lumbopelvic rhythm (Esola et al., 1996; López-Miñarro and Alacid, 2009) and low back pain (Biering-Sorensen, 1984; Mierau et al., 1989). Additionally, individuals with shortened hamstring muscles present gait limitations, increased risk of falls, and susceptibility to musculoskeletal injuries (Erkula et al., 2002; Jones et al., 1998).

Nowadays different kinds of tests are used to assess hamstring extensibility. Flexibility is typically characterized by the maximum range of motion in a joint or series of joints (McHugh et al., 1998). Thus, angular tests that specifically measure hip flexion with the knee extended (straight leg raise test), or the range of knee extension with the hip flexed to 90 degrees (knee extension or popliteal angle test), have been widely considered the criterion measures of hamstring extensibility (e.g., Ayala et al., 2011; Hartman and Looney, 2003; López-Miñarro and Rodríguez-García, 2010c). Nevertheless, due to the necessity of sophisticated instruments, qualified technicians, and time constraints, the use of these angular tests seem to be limited in several settings such as in a school context or large scale studies (Castro-Piñero et al., 2009b).

Unlike the angular tests, lineal tests have a simple procedure, are easy to administer, require-minimal skills training for their application, and the equipment necessary to perform them is very affordable (Castro-Piñero et al., 2009b; López Miñarro et al., 2008c). Sit-and-reach (SR) tests in which a fingertips-to-tangent feet distance is measured are probably the most widely used lineal measures of flexibility (Holt et al., 1999; Castro-Piñero et al., 2009a). However, as the SR is a test which involves the movement of the whole body, it has been suggested that the position of the fingertips does not give valid information about hamstring extensibility (Hoeger et al., 1990). The main factors that seem to affect the validity of SR tests to estimate hamstring extensibility are the differences in length proportion between the upper and lower limbs (Hoeger et al., 1990), the position of the head (Smith and Miller, 1985) and the position of the ankles (Kawano et al., 2010; Liemohn et al., 1997). In addition, recent studies have also found that the levels of hamstring extensibility influence the criterion-related validity of SR tests (López-Miñarro et al., 2011; López-Miñarro and Rodríguez-García, 2010c).

The choice of a flexibility test must be based on its functionality and validity (López-Miñarro, 2010). Although the angular tests have the advantage of being the criterion measure to assess flexibility, due to several practical reasons they have the disadvantage of having a limited use in several settings (Castro-Piñero et al., 2009b). In these settings, as the SR tests have the advantage of allowing for an evaluation in a short amount of time with

Received: 21 June 2013 / Accepted: 22 August 2013 / First Available (online): 08 October 2013 / Published (online): 20 January 2014

minimal skills and instruments, potentially they could be a useful alternative to estimate flexibility. Nevertheless, as in the application of any fitness field test, the SR tests' results are a simple estimation and, therefore, the evaluators must be aware of validity coefficients in order to interpret the scores of these tests correctly. Unfortunately, the studies examining criterion-related validity of SR tests for estimating hamstring and lumbar extensibility have shown inconclusive results (Baltaci et al., 2003; Hui and Yuen, 2000; Hui et al., 1999; Jones et al., 1998).

Each primary study that is published about criterion-related validity of the SR tests only constitutes as a single piece of a constantly growing body of evidence (Cooper et al., 2009). For example, in some studies the correlation coefficient is statistically significant, while in others a statistically significant association is not found. In some cases the strength of the association is quite high, while low in others. To make sense of the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses (Cooper et al., 2009; Hunter and Schmidt, 2004; Lipsey and Wilson, 2001). Hence, the meta-analyses remain a useful tool for the evaluation of evidence (Flather et al., 1997), forming a critical process for theory development in science (Hunter and Schmidt, 2004).

Unfortunately, to our knowledge there are not any meta-analyses addressing the criterion-related validity of SR tests. Beyond the simple but important function of describing and summarizing the scientific findings of a research area, the main contribution of a meta-analysis is to estimate as accurately as possible the population parameters (Hunter and Schmidt, 2004). Therefore, the results of a meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies. Finally, the meta-analyses permit us to examine today´s lack of knowledge in a specific area and to guide scientists in future research (Cooper et al., 2009).

Consequently, the main purpose of the present meta-analysis was to examine the scientific literature on criterion-related validity of SR tests for estimating hamstring and lumbar extensibility in apparently healthy individuals. More specifically, the objectives of this study were: (a) to describe and summarize the up-to-date scientific findings of criterion-related validity of SR tests for estimating hamstring and lumbar extensibility; (b) to estimate and compare the overall population mean of the criterion-related validity coefficients of each SR test for estimating hamstring and lumbar extensibility; and (c) to examine the influence of some study features (sex of the participants, age of participants, and level of hamstring extensibility) in criterion-related validity coefficients of SR tests.

Methods

Search strategy

The following seven electronic databases were searched from their inception through December 2012: SportDiscus, Scopus, Medline, Pubmed, Web of Science, ERIC, and Dissertations & Theses Database. The search terms used were based on two concepts. Concept one included terms for the SR test (sit and reach) and concept two included terms for validity (validity, related, relationship, correlation, comparison, hip, hamstring, flexibility, ROM, range of motion, range of movement, straight leg raise, knee extension, popliteal angle, lumbar, back, Macrae and Wright, Macrae & Wright, Schober, radiography, goniometer, and inclinometer). The terms of the same concept were combined together with the Boolean operator "OR" and then the two concepts were combined using the Boolean operator ''AND'' (Benito Peinado et al., 2007). The keywords that consisted of more than one word were enclosed in quotes. In addition, the reference lists of all included papers were manually searched.

Selection criteria

The selection criteria to identify studies that examined the criterion-related validity of SR tests for estimating hamstring and/or lumbar extensibility were: (a) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (b) studies with SR tests that yielded the values of the maximum reach of the fingertips; and (c) studies in which hamstring and/or lumbar extensibility criterion measurements used are widely accepted in the scientific literature (e.g., straight leg raise or knee extension tests for hamstring extensibility and Macrae & Wright or inclinometer methods for lumbar extensibility). In addition to papers, master/doctoral dissertations and conference proceedings were also accepted. No language or publication date restrictions were imposed.

Coding studies

For this meta-analysis, data were collected from studies that reported relationships between SR tests and hamstring and/or lumbar extensibility criterion measures with apparently healthy participants of any age. From each selected study the following data were coded: Study identity number, sample size (n), sex of participants (1 = males, 2 = females), age of participants (1 = children, $<$ 18 years; 2 = adults, $\geq$ 18 years), SR test protocol (1 = Classic SR, 2 = Modified SR, 3 = Back-saver SR, 4 = Modified back-saver SR, 5 = V SR, 6 = Modified V SR, 7 = Unilateral SR, 8 = Chair SR), criterion-related validity value (Pearson´s r correlation coefficient), reliability of SR test (intraclass correlation coefficient), reliability of hamstring and/or lumbar extensibility criterion measures (intraclass correlation coefficient), and the average score of hamstring extensibility criterion measure. Because identification of study features is usually explicitly stated in each of the primary articles, the use of more than one rather was deemed unnecessary.

In addition, although various protocols for evaluating quality of single studies have been described, there is no widespread agreement on the validity of this type of evaluation approach. Thus, rejecting certain single studies and accepting others for inclusion in a meta-analysis on the basis of a quality score remains a controversial procedure (Flather et al., 1997). Hence, according to Flather et al. (1997), our approach has been to ensure that the design has not been flawed (e.g., conducted by scientifically evidenced criterion measures), and that there has been a complete reporting of relevant outcomes. For a study to be included in this meta-analysis, sample size, SR test protocol, hamstring and/or lumbar criterion measures and Pearson's *r* were considered to be critical. In the event that the authors mixed subgroups of a study feature (e.g., males mixed with females), failed to identify a study feature (e.g., criterion measure or reliability scores) or were ambiguous (e.g., hamstring extensibility scores around 80º shown graphically) the data was omitted. When in the same study data for males and females were expressed both separately and together, only the separate data were coded. When in the same study data were expressed for both legs separately or for two different days from the same sample (i.e., such as in Mier, 2011), the average value of the coefficients was coded.

Finally, in the event that included studies used multiple validity coefficients for hamstring and/or lumbar extensibility, only the data relative to one criterion measure of each muscle group was coded. Regarding hamstring extensibility, all studies reported correlation values with the straight leg raise test, while only in a few articles the values with the knee extension test was also stated (Davis et al., 2008; García, 1995; Harman and Looney, 2003). Therefore, in order to avoid moderator effects issues by criterion measure test, only the correlation values of the straight leg raise test were coded. As regards lumbar extensibility, only Hartman and Looney (2003) performed more than one criterion measure test (Single inclinometer and Macrae & Wright methods). Due to the fact that the Macrae & Wright method has been used the most widely, the results with this test were coded.

Data analyses

In the present study, Pearson´s zero-order correlation coefficient (*r*) was considered the unit of measure as an indication of criterion-related validity of SR tests, which represents the strength of associations between the estimates of SR tests and the criterion measures. Because several studies reported criterion-related validity results of different SR test protocols from the same sample, *r* values were extracted separately for each SR test to avoid dependency issues in the meta-analysis (Cooper et al., 2009). Similarly, criterion-related validity values were extracted separately for hamstring and lumbar extensibility. However, if a single study reported more than one *r* value within the same SR test protocol, but from different subsamples (e.g., males and females), we assumed each *r* value from different subsamples to be independent from each other and included them in a single meta-analysis (Lipsey and Wilson, 2001).

Publication bias: In addition to the followed search strategy and selection criteria to avoid availability bias, an examination of the selected studies was carried out to avoid a potential duplication of information retrieved. Since some selected studies had full or partial duplicated information, these particular *r* correlations values were not analyzed in the meta-analyses. Furthermore, before computing correlations, several exploratory analyses were also conducted to detect the presence of publication bias. Firstly, a file drawer analysis based on effect size was performed to estimate the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean effect size (r_p) down to the small mean *r* value (Rosenthal, 1979). Depending on the results of the file drawer analysis, we had to conclude if it was likely that there would be this particular number of "lost" studies to reduce the actual *r* to a small value. According to Cohen´s guidelines (1992), the correlation coefficient was interpreted as small when $r < 0.30$.

Secondly, according to Light and Pillemer (1984), the scatter plots of correlations coefficients against sample size for each SR test protocol related to both hamstring and lumbar extensibility were analyzed. According to this graphic method, in the absence of publication bias, the resulting figure should take the form of an inverted funnel. However, based on the statistical significance of the studies, if there is publication bias the small-sample studies reporting small *r* values will be disproportionately absent because they are the studies that will fail to attain statistical significance. Finally, with the objective of quantifying the outcomes of the scatter plots, as suggested by Begg and Mazumdar (1994), a Spearman´s rank order correlation between *r* values and sample size was calculated. In the presence of publication bias, this correlation should be statistically significant negative due to the absence of small-sample studies in the lower left hand corner.

Computation of correlations: The Hunter-Schmidt´s psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of SR tests (Hunter and Schmidt, 2004). This approach estimates the population correlation by individually correcting the observed correlations due to various artefacts such as sampling error and measurement error. First, the "bare-bone" mean *r* (r_c), corrected for only sampling error, was calculated by weighting each *r* with the respective sample size when aggregating them into r_c. Then, we calculated the corrected mean *r* at the population level (r_p) that was unaffected by both sampling error and measurement error. The resulting mean correlation corrected for sampling error and measurement error is offered as the best estimate of the population parameter. In order to correct the measurement errors, the reliability coefficients (intraclass correlation coefficients) of each individual SR and criterion measure tests were used. Because the reliability coefficients were not available for all of the included studies, the unknown reliability values were previously estimated for each test. The median of the all reported reliability coefficients for each SR test protocol and criterion measure test was used. Finally, the 95% confidence intervals of r_p (95% CI) were calculated.

Moderator analysis: In the present meta-analysis, due to the low number of *r* values found, partially hierarchical analyses of moderator variables were carried out. According to Hunter and Schmidt (2004), to determine the presence of moderator effects which may affect overall criterion-related validity of SR tests (r_p), three different criteria were simultaneously examined: (a) the percentage of variance accounted for by statistical artefacts is less than 75% of the observed variance in r_p; (b) the *Q* homogeneity statistic is statistically significant ($p <$

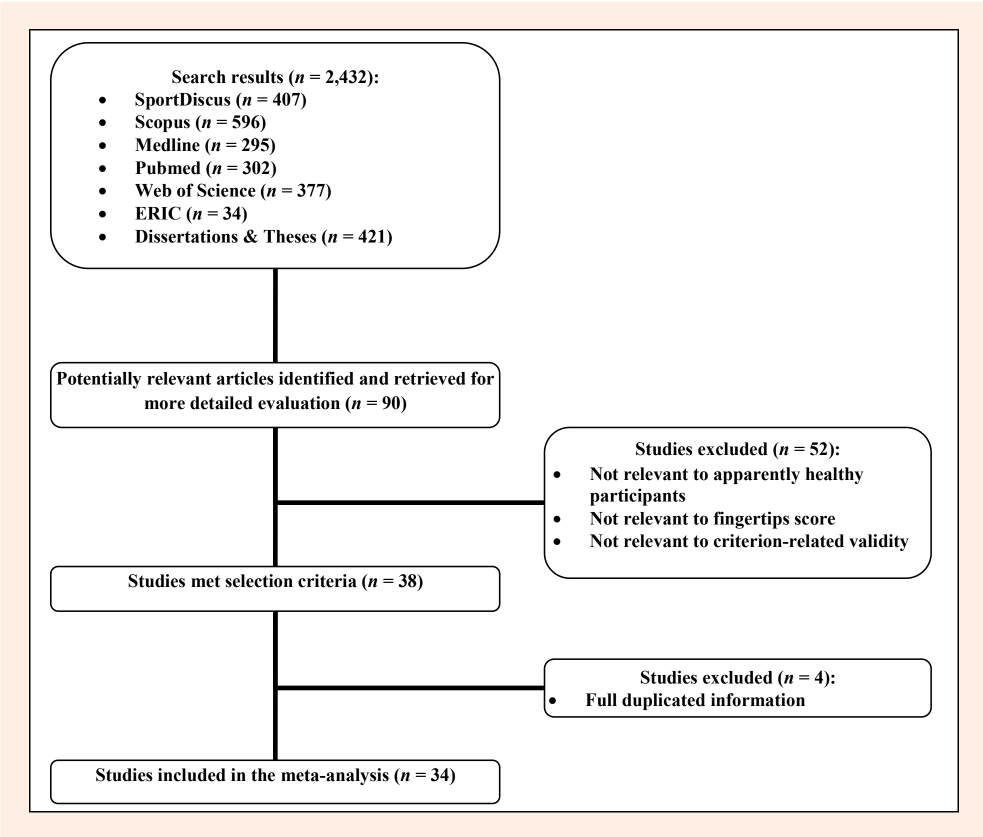


Figure 1. Flow chart of studies selection process.

0.05); and (c) the 95% credibly interval (95% CV) is relatively large or includes the value zero. If at least one of the three criteria were met, we concluded that the results could be affected by moderator effects. In case of the presence of moderator effects, criterion-related validity values of each SR test were analyzed separately by: (a) sex of participants (i.e., male and female); (b) age of participants (i.e., children and adults); and (c) level of hamstring extensibility (i.e., low average level, < 80º, and high average level, ≥ 80º) (Kendall et al., 2005).

Results

Study description

Figure 1 shows a flow chart of the study selection process. Of the 2,432 literature search results, 90 potentially relevant publications were identified and retrieved for a more detailed evaluation. Finally, due to duplication issues, of the 38 studies that met the inclusion criteria, only 34 studies were included in the present meta-analysis. Apart from a few studies retrieved which were carried out with apparently non-healthy participants or lineal tests that did not yield the values of the maximum reach of the fingertips, other studies (or *r* values) were not included either in the present meta-analysis because they examined the relationship between the SR test and the pelvic tilt scores (e.g., Davis et al., 2008; Kawano et al., 2010; López-Miñarro, 2010; Rodríguez-García et al., 2008). The pelvic tilt is measured by the inclination angle of the sacrum with regard to the horizontal line at the point of maximal forward reach on the SR test. Therefore, although the pelvis position is influenced by the hamstring extensibility, its measure must be considered as an estimation of hamstring extensibility (indirect measure), and not as a criterion measure to determinate it (direct measure) such as the straight leg raise or knee extension tests (Santonja Medina et al., 1995). However, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman and Looney, 2003) (see strengths and limitations section).

Table 1 presents a summary of studies of criterion-related validity of SR tests for estimating hamstring and lumbar extensibility. Regarding the criterion-related validity for estimating hamstring extensibility, a total of 99 *r* values across eight SR test protocols were retrieved, ranging from three values in the Chair SR and Modified V SR tests to 47 values in the Classic SR test. Total sample sizes for each SR test ranged from 182 in the Chair SR test to 3,481 in the Classic SR test. The individual criterion-related validity correlation coefficients of SR tests for estimating hamstring extensibility ranged from 0.19 to 0.93. Regarding criterion-related validity for estimating lumbar extensibility, a total of 51 *r* values across seven SR test protocols were retrieved, ranging from two values in the Unilateral SR test to 21 values in Classic SR test. Studies examining the criterion-related validity of the Chair SR test for estimating lumbar extensibility were not found. Total sample sizes for each SR test ranged from 158 in the Unilateral SR test to 1,762 in Classic SR test. The individual criterion-related validity correlation coefficients of SR tests for estimating lumbar extensibility ranged from 0.00 to 0.60.

Table 1. **Summary of studies of criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility.**

Reference	Sample	*n*	Age (yrs)	Test	Hamstring extensibility			Lumbar extensibility		
					Criter	♂ (*r*)	♀ (*r*)	Criter	♂ (*r*)	♀ (*r*)
Ayala et al. (2011)	Professional futsal players	♂=55 ♀=48	26.0 (4.5) 23.0 (5.3)	CSR MSR BSSR	PSLR PSLR PSLR	.62* .76* .47	.93* .73* .91*			
Ayala et al. (2012)	Recreationally active university students	♂=156 ♀=87	21.3 (2.5) 20.7 (1.6)	CSR	PSLR	.79*				
Baker (1985)	High and Middle school students	♀=100	14.1 (.8)	CSR	PSLR		**.64***	MWM		**.28***
Baltaci et al. (2003)	University students	♀=102	22.0 (1.0)	CSR BSSR CHSR	PSLR PSLR PSLR		.63*-.53* .37*-.44* .22*-.16			
Book (1989)	Public and private schools students	♂=203 ♀=255	9-18	CSR	PSLR	.65*	.81*	MWM	.33*	.30*
Bozic et al. (2010)	Physically active sport and PE white students	♂=84	21.3 (2.6)	CSR	PSLR	.63*				
Castro-Piñero et al. (2009b)	Caucasian children and adolescents	♂=29 ♀=27 ♂=16 ♀=15	6-12 13-17	CSR MSR CSR MSR	PSLR PSLR PSLR PSLR	.38* .34* .38* .26				
Chung and Yuen (1999)	University students	♂=52	20.7 (1.3)	CSR MSR	PSLR PSLR	.77* .71*		MWM MWM	.23 .24	
Davis et al. (2008)	University students	♂=42 ♀=39	23.6 (4.1) 24.1 (4.3)	CSR CSR	PSLR PKE	.65* .57*				
García (1995)	High school students	♂=54 ♀=55	15-18	CSR CSR	PSLR AKE	.64* .63*	.67* .61*	MWM	.32*	.28*
Hartman and Looney (2003)	Schoolchildren	♂=85 ♀=88	6-12	CSR CSR BSSR BSSR	PSLR AKE PSLR AKE	.67*-.66* .40*-.40* .69*-.67* .50*-.47*	.47*-.49* .52*-.54* .42*-.48* .54*-.57*	MWM SIM MWM SIM	.05 .29* .00-.07 .26*-.28*	.07 .16 .06-.06 .10-.10
Hui et al. (1999)	University students	♂=62 ♀=96	21.1 (3.1) 20.6 (2.1)	CSR MSR BSSR MBSSR VSR MVSR	PSLR PSLR PSLR PSLR PSLR PSLR	.48*-.47* .45*-.45* .46*-.44* .44*-.45* .58*-.63* .57*-.62*	.46*-.53* .41*-.47* .39*-.50* .35*-.47* .44*-.52* .46*-.51*	MWM MWM MWM MWM MWM MWM	.27* .24 .24-.27* .17-.20 .42* .38*	.24* .22* .18-.15 .22*-.22* .24* .28*
Hui and Yuen (2000)	University students	♂=62 ♀=96	21.1 (3.1) 20.6 (2.1)	CSR BSSR VSR USR	PSLR PSLR PSLR PSLR	**.48*-.47*** **.46*-.44*** **.58*-.63*** .61*-.67*	**.46*-.53*** **.39*-.50*** **.44*-.52*** .50*-.54*	MWM MWM MWM MWM	**.27*** **.24-.27*** **.42*** .47*-.47*	**.24*** **.18-.15** **.24*** .26*-.23*
Jackson and Baker (1986)	School PE students	♀=100	14.1 (.8)	CSR	PSLR		.64*	MWM		.28*
Jackson and Langford (1989)	?	♂=52 ♀=52	20-45	CSR	PSLR	.89*	.70*	MWM	.59*	.12
Jones et al. (1998)	Exercise classes at a retirement community	♂=32 ♀=48	74.5 (5.7) 74.0 (6.7)	CSR BSSR CHSR	PSLR PSLR PSLR	.74* .70* .76*	.71* .71* .81*			
Kanbur et al. (2005)	Non-prepubertal/ non-regularly exercised boys	♂=69	13-14	CSR	PSLR	.64*-.65*				
Langford (1987)	?	♂=52 ♀=52	20-45	CSR	PSLR	**.89***	**.70***	MWM	**.60***	**.12**
Lemmink et al. (2003)	Independently living people over 55	♂=49 ♀=71	67.7 (7.5) 65.6 (8.6)	CSR MSR	PSLR PSLR	.74* .54*	.57* .57*	AAOS AAOS	.13 .21	.31* .26*
Liemohn et al. (1994)	University students	♂=20 ♀=20	24.0 (4.6) 25.1 (6.3)	CSR BSSR	PSLR PSLR	.72* .76*	.70* .70*	SIM SIM	.29 .32	.40 .38

This table includes all studies that met selection criteria, however, full or partial information was not included in the meta-analysis (in bold) due to duplication issues; ♂, males; ♀, females; ?, information unavailable; Criter: Criterion, CSR, Classic sit-and-reach test; MSR, Modified sit-and-reach test; BSSR, Back-saver sit-and-reach test; MBSSR, Modified back-saver sit-and-reach test; VSR, V sit-and-reach test; MVSR, Modified v sit-and-reach test; USR, Unilateral sit-and-reach test; CHSR, Chair sit-and-reach test; PSLR, Passive straight leg raise test; ASLR, Active straight leg raise test; PKE, Passive knee extension test; AKE, Active knee extension test; SMM, Spinal Mouse method; SIM, Single Inclinometer method; MWM, Macrae & Wright method; AAOSM, American Academy of Orthopedic Surgeons method;, Pearson´s *r* for the left and right leg, respectively. * Pearson´s *r* statistically significant at $p < 0.05$

Table 1. **Continued.**

Reference	Sample	*n*	Age (yrs)	Test	Hamstring extensibility			Lumbar extensibility		
					Criter	♂ (*r*)	♀ (*r*)	Criter	♂ (*r*)	♀ (*r*)
López-Miñarro et al. (2008)	University students	♂=102	22.9 (3.2)	CSR	PSLR	.56*-.59*	.72*-.74*	SIM	.32*	.14
		♀=96	23.2 (4.5)	VSR	PSLR	.53*-.55*	.63*-.65*	SIM	.33*	.29*
López Miñarro et al. (2008a)	Canoeists	♂=44	13.3 (.6)	CSR	PSLR	.77*-.73*	.74*-.81*			
		♀=22								
López Miñarro et al. (2008b)	Canoeists	?=66	13.3 (.6)	CSR	PSLR	**.70*-.68***				
López Miñarro et al. (2008c)	?	♂=120	22.8 (3.1)	CSR	PSLR	**.56*-.59***	**.72*-.74***			
		♀=100	23.1 (4.6)	BSSR	PSLR	.51*-.50*	.68*-.68*			
				USR	PSLR	.54*-.58*	.73*-.75*			
López-Miñarro et al. (2009)	University students	♂=76	23.5 (4.0)	CSR	PSLR	.56*-.59*	.75*-.73*			
		♀=67	23.9 (5.4)	BSSR	PSLR	.53*-.51*	.70*-.66*			
López-Miñarro et al. (2010a)	?	♂=73	23.0 (3.5)	CSR	PSLR	.44*-.48*	.75*-.73*			
		♀=71	23.1 (4.3)	MSR	PSLR	.28*-.32*	.63*-.64*			
López-Miñarro et al. (2010b)	University students	♂=130	22.9 (3.2)	CSR	PSLR	**.56*-.59***	**.72*-.74***			
		♀=110	23.2 (4.5)	MSR	PSLR	.41*-.45*	.62*-.63*			
				BSSR	PSLR	**.51*-.49***	**.68*-.68***			
				VSR	PSLR	**.53*-.55***	**.63*-.65***			
López-Miñarro and Rodríguez-García (2010c)	Recreationally active university students: Low and normal flexibility	♂=120	22.9 (3.6)	CSR	PSLR	.31*-.41*				
		♂=120		CSR	PSLR	.61*-.55*				
López-Miñarro et al. (2011)	Older women: Low, moderate and high flexibility	♀=36	65.3 (9.1)	CSR	PSLR		.43*-.41*			
		♀=35					.54*-.57*			
		♀=35					.73*-.70*			
López-Miñarro et al. (2012)	Canoeists	♂=51	17.5 (6.3)	CSR	PSLR	.67*-.66*				
	Kayakers	♂=60				.59*-.59*				
Mier (2011)	Physically active adults	♂=30	25.0 (9.3)	CSR	PSLR	.64*/.66*	.79*/.81*			
		♀=30	23.7 (7.9)	CSR	PSLR					
Minkler and Patterson (1994)	Regular PE activity classes practitioners (mainly Caucasians)	♂=48	24.3 (4.7)	MSR	PSLR	.75*	.66*	MWM	.40*	.25
		♀=51	21.5 (3.8)							
Miyazaki et al. (2010)	Community-dwelling elderly	♂=42	72.6 (6.9)	CSR	PSLR	.60*		SMM	.18	
		♀=119								
Orloff (1988)	Gymnasium practitioners	♂=47	19-54	CSR	ASLR	.52*				
		♀=28								
Patterson et al. (1996)	Middle school students (various ethnic origins)	♂=40	13.0 (.9)	BSSR	PSLR	.72*-.68*	.51*-.52*	MWM	.15-.10	.17-.25
		♀=44	12.7 (.8)							
Rodríguez-García et al. (2008)	Fit sports activities practitioners	♂=125	22.9 (3.2)	CSR	PSLR	**.56*-.59***	**.72*-.74***	SIM	**.32***	**.14**
		♀=118	23.2 (4.5)							
Simoneau (1998)	Physically active university students	♀=34	20.3 (.9)	CSR	PSLR		.78*	MWM		.26
Yuen and Hui (1998)	University students	♂=19	?	CSR	PSLR	.52*-.57*		MWM	.18	
		♀=36		MSR	PSLR	.49*-.55*		MWM	.27	
				BSSR	PSLR	.46*-.44*		MWM	.16-.12	
				MBSSR	PSLR	.39*-.52*		MWM	.21-.17	
				VSR	PSLR	.54*-.60*		MWM	.19	
				MVSR	PSLR	.58*-.61*		MWM	.25	

Note. This table includes all studies that met selection criteria, however, full or partial information was not included in the meta-analysis (in bold) due to duplication issues; ♂, males; ♀, females; ?, information unavailable; Criter, Criterion, CSR, Classic sit-and-reach test; MSR, Modified sit-and-reach test; BSSR, Back-saver sit-and-reach test; MBSSR, Modified back-saver sit-and-reach test; VSR, V sit-and-reach test; MVSR, Modified v sit-and-reach test; USR, Unilateral sit-and-reach test; CHSR, Chair sit-and-reach test; PSLR, Passive straight leg raise test; ASLR, Active straight leg raise test; PKE, Passive knee extension test; AKE, Active knee extension test; SMM, Spinal Mouse method; SIM, Single Inclinometer method; MWM, Macrae & Wright method; AAOSM, American Academy of Orthopedic Surgeons method; Pearson´s *r* for the left and right leg, respectively.
* Pearson´s *r* statistically significant at $p < 0.05$

Publication bias

Due to some studies having fully or partially duplicated information, these *r* coefficients values were not analyzed in the present meta-analyses despite the fact that these studies met the selection criteria. For example, Baker (1985) and Langford´s (1987) doctoral dissertations were not included because the data were published later in a journal (although in Langford´s works there was a little difference in one *r* value, it was simply considered a typo because the other data were equal) (Jackson and Baker, 1986; Jackson and Langford, 1989). López Miñarro´s et al. (2008b) study information (males mixed with females) were not computed because the same data were also published with males and females separately (López Miñarro et al., 2008a). Additionally, full or partial information from a few studies of the same authors, sample characteristics, and correlation results was not included either due to duplication issues (Hui and Yuen, 2000; López Miñarro et al., 2008c; López-Miñarro et al., 2010b; Rodríguez-García et al., 2008). Pearson´s r correlation values of selected studies that were excluded for meta-analysis are indicated (in bold) in Table 1.

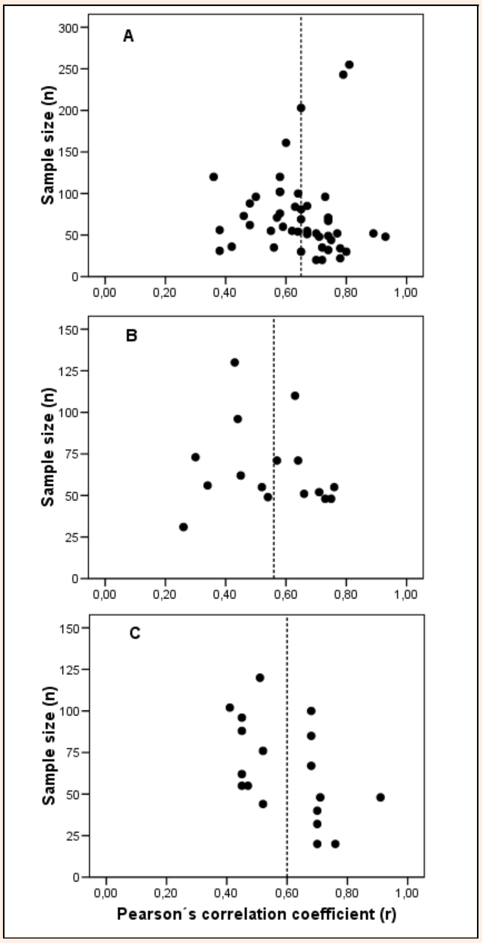


Figure 2. Scatter plots of sample size and criterion-related validity coefficients (*r*) for estimating hamstring extensibility: (a) Classic sit-and-reach; (b) Modified sit-and-reach; and (c) Back-saver sit-and-reach. Dashed line represents median values of validity coefficients.

Afterward, several exploratory analyses were conducted to detect the presence of publication bias. Regarding hamstring extensibility, the results of the file drawer analyses are based on effect size for estimating the number of unlocated studies averaging null results ($r = 0$) that would have to exist to bring the mean r_p down to 0.29. These results are shown in the following lines (in parenthesis the unlocated/located percentage): 63 for the Classic SR (134%), 15 for the Modified SR (94%), 19 for the Back-saver SR (106%), 2 for the Modified back-saver SR (67%), 6 for the V SR (120%), 4 for the Modified V SR (133%), 5 for the Unilateral SR (125%), and 3 for the Chair SR (100%). Although we are aware that there is not a large number of "lost" studies for the Modified back-saver SR, V SR, Modified V SR, Unilateral SR, and Chair SR, the percentage of unlocated/located studies was unlikely (67-133%). Hence, we concluded that it was unlikely that there would be this particular number of "lost" studies for each SR test protocol. On the other hand, regarding the lumbar extensibility, the file drawer analyses were not calculated because the actual r values were small.

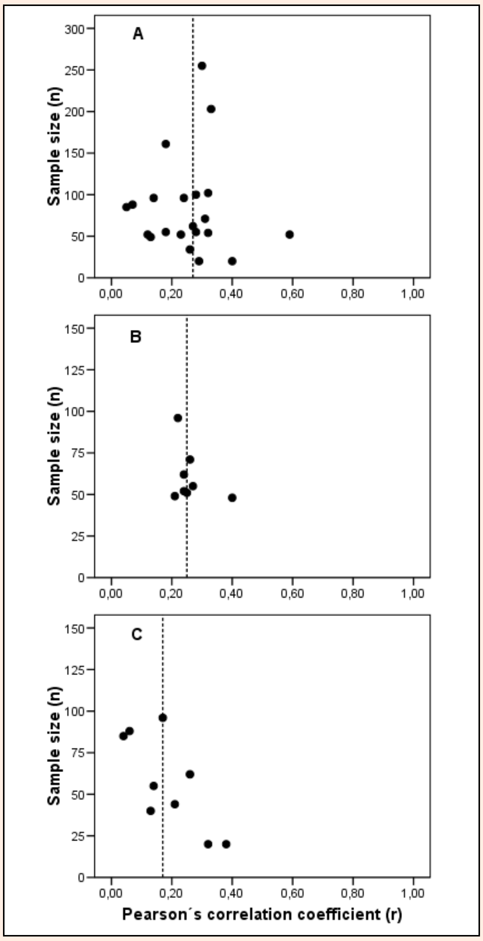


Figure 3. Scatter plots of sample size and criterion-related validity coefficients (r) for estimating lumbar extensibility: (a) Classic sit-and-reach; (b) Modified sit-and-reach; and (c) Back-saver sit-and-reach. Dashed line represents median values of validity coefficients.

Figures 2 and 3 show the scatter plots of sample size against criterion-related validity coefficients for estimating hamstring and lumbar extensibility, respectively. Due to the low number of r values for the most SR test protocols (2-5 r values), only the scatter plots for the Classic SR, Modified SR, and Back-saver SR tests were examined. According to this graphic method, the figures suggested that there was an absence of publication bias for the Classic SR and Modified SR tests. However, the two scatter plots of the Back-saver SR test suggested the presence of publication bias, because of the absence of r values in the lower left hand corner of the inverted funnel plot. In this line, for the Back-saver SR test, the results of Spearman´s rank order correlation between r values and sample size showed a statistically significant negative correlation for estimating hamstring extensibility (r = -

0.66, p = 0.003) and marginally significant for lumbar extensibility (r = -0.61, p = 0.081). Nevertheless, for the Classic SR and Modified SR tests the results did not show a statistically significant correlation for either estimating hamstring (Classic SR, r = -0.29, p = 0.050; Modified SR, r = -0.33, p = 0.207) or lumbar extensibility (Classic SR, r = -0.02, p = 0.935; Modified SR, r = -0.22, p = 0.608). Finally, although we aware that the results for the Classic SR test for estimating hamstring extensibility were marginally significant, the r value was considerably lower than for the Back-saver SR test.

Criterion-related validity

Table 2 reports the number of studies (K), the cumulative number of r values (n), the total sample size accumulated (N), the overall weighted mean of r corrected for sampling error only (r_c), the overall weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI for overall criterion-related validity correlation coefficients (r_p) separately for estimating hamstring and lumbar extensibility across each SR test protocol. In addition, to detect the presence of moderator effects which may affect overall criterion-related validity of SR tests, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the Q homogeneity statistic were calculated.

Hamstring extensibility: The overall results showed that all SR test protocols had a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility (r_p range = 0.46-0.67) in which all 95% CI did not include the value zero. For five of the eight SR test protocols, the percentage of variance accounted for by statistical artefacts was less than 75%, the Q homogeneity statistic was statistically significant (p < 0.05), and the 95% CV was relatively large. Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study.

Lumbar extensibility: The overall results showed that all SR test protocols had a low mean correlation coefficient of criterion-related validity for estimating lumbar extensibility (r_p range = 0.16-0.35) in which, the 95% CI of the Back-saver SR and the Modified back-saver SR tests included the value zero. Furthermore, studies addressing the criterion-related validity of the Chair SR test for estimating lumbar extensibility were not found. Finally, since none of the three criteria were met in the seven SR test protocols, moderator analyses were not conducted for lumbar extensibility.

Moderator analyses

Table 3 reports the results of moderator analyses to examine the effects of the sex of the participants (i.e., male and female), the age of participants (i.e., children and adults), and the level of hamstring extensibility (i.e., low average level, < 80º, and high average level, ≥ 80º) on overall criterion-related validity correlation coefficients for estimating hamstring extensibility for each SR test protocol potentially affected by moderator effects (i.e., the Classic SR, Modified SR, Back-saver SR, Unilateral SR, and Chair SR). Collectively, slight differences in r_p values were detected in different categories of included moderators across the analyzed SR tests.

Gender of participants: The results showed that all SR test protocols had a moderate-to-high mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for males (r_p range = 0.55-0.83) and moderate for females (r_p range = 0.41-0.70) in which all 95% CI did not include the value zero. There was a tendency of the mean correlation coefficient being slightly greater for females than for males on each SR test, except for the Chair SR test where the opposite results were found. However, we have to be aware that, except for the Chair SR test, all the 95% CI of mean correlation coefficients were overlapped. Moreover, we should also be cautious because the low numbers of r values over the analyses were supported. Additionally, according to moderator analysis criteria, at least one of

Table 2. Results of meta-analyses for overall criterion-related validity correlation coefficients across sit-and-reach test protocols.

Sit-and-reach test	K	n	N	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	Q statistic
Hamstring extensibility									
Classic sit-and-reach	28	47	3,481	.65	.67	.55, .80	.44, .91	22.55	208.39*
Modified sit-and-reach	9	16	1,058	.54	.56	.39, .73	.32, .80	33.14	48.28*
Back-saver sit-and-reach	10	18	1,158	.57	.59	.43, .75	.38, .80	36.65	49.12*
Modified back-saver sit-and-reach	2	3	213	.44	.46	.28, .65	.46, .46	100.00	.18
V sit-and-reach	3	5	411	.56	.60	.46, .74	.60, .60	100.00	3.23
Modified V sit-and-reach	2	3	213	.55	.59	.44, .74	.59, .59	100.00	1.14
Unilateral sit-and-reach	2	4	378	.61	.64	.52, .76	.51, .76	47.43	8.43*
Chair sit-and-reach	2	3	182	.45	.49	.29, .68	-.11, 1.00	9.50	31.57*
Lumbar extensibility									
Classic sit-and-reach	13	21	1,762	.25	.26	.05, .46	.19, .32	91.52	22.95
Modified sit-and-reach	5	8	484	.26	.26	.03, .50	.26, .26	100.00	1.39
Back-saver sit-and-reach	5	9	510	.15	.16	-.10, .41	.16, .16	100.00	4.72
Modified back-saver sit-and-reach	2	3	213	.20	.21	-.01, .44	.21, .21	100.00	.07
V sit-and-reach	3	5	411	.30	.31	.11, .51	.31, .31	100.00	2.42
Modified V sit-and-reach	2	3	213	.30	.32	.11, .53	.32, .32	100.00	.63
Unilateral sit-and-reach	1	2	158	.34	.35	.15, .54	.26, .43	83.73	2.39
Chair sit-and-reach	-	-	-	-	-	-	-	-	-

Note. K, number of studies; n, number of rs; N, total sample size; r_c, overall weighted mean of r corrected for sampling error only; r_p, overall weighted mean of r corrected for sampling error and measurement error; [a] 95% confidence interval; [b] 95% credibly interval [c] Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of sit-and-reach tests. * p < 0.05

Table 3. Results of moderator analyses for criterion-related validity correlation coefficients for estimating hamstring extensibility across all sit-and-reach test protocols potentially affected by moderator effects†

Moderator	Effect	*K*	*n*	*N*	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	*Q* statistic
Gender of participants										
Classic sit-and-reach	Males	19	21	1,493	.62	.64	.50, .78	.46, .82	38.28	54.86*
	Females	18	20	1,361	.68	.70	.58, .82	.49, .91	25.06	79.80*
Modified sit-and-reach	Males	7	7	469	.53	.55	.38, .71	.26, .83	26.80	59.70*
	Females	6	6	447	.60	.62	.48, .76	.53, .72	68.72	23.28*
Back-saver sit-and-reach	Males	8	8	490	.57	.59	.42, .75	.48, .69	71.70	11.16
	Females	9	9	613	.58	.60	.45, .75	.33, .87	24.05	37.42*
Unilateral sit-and-reach	Males	2	2	182	.59	.61	.48, .74	.61, .61	100.00	.72
	Females	2	2	196	.63	.66	.55, .77	.47, .85	25.66	7.80*
Chair sit-and-reach	Males	1	1	32	.76	.83	.71, .94	.83, .83	100.00	.00
	Females	2	2	150	.39	.41	.22, .60	-.16, .99	9.79	20.42
Age of participants										
Classic sit-and-reach	Children	8	14	1,173	.66	.67	.55, .79	.47, .87	26.28	53.27*
	Adults	20	33	2,308	.64	.68	.55, .80	.43, .92	21.22	155.50*
Modified sit-and-reach	Children	1	2	87	.31	.32	.05, .59	.32, .32	100.00	1.34
	Adults	8	14	971	.56	.58	.42, .74	.37, .79	35.13	45.55*
Back-saver sit-and-reach	Children	2	4	257	.58	.59	.43, .75	.45, .74	55.88	7.16
	Adults	8	14	901	.56	.59	.43, .75	.36, .82	33.38	41.94*
Unilateral sit-and-reach	Children	-	-	-	-	-	-	-	-	-
	Adults	2	4	378	.61	.64	.52, .76	.51, .76	47.43	8.43*
Chair sit-and-reach	Children	-	-	-	-	-	-		-	-
	Adults	2	3	182	.45	.49	.29, .68	-.11, 1.00	9.50	31.57*
Level of hamstring extensibility										
Classic sit-and-reach	Low	15	16	1,129	.60	.63	.48, .77	.41, .84	30.79	51.97*
	High	19	25	1,984	.67	.70	.59, .81	.46, .94	18.39	135.91*
Modified sit-and-reach	Low	4	5	355	.51	.53	.36, .70	.24, .82	24.87	64.35*
	High	7	10	648	.55	.58	.41, .74	.36, .79	37.42	42.75*
Back-saver sit-and-reach	Low	5	6	433	.54	.57	.41, .72	.31, .82	27.37	21.92*
	High	7	11	670	.59	.61	.45, .77	.44, .79	45.34	24.26*
Unilateral sit-and-reach	Low	1	1	120	.56	.58	.47, .70	.58, .58	100.00	.00
	High	2	3	258	.63	.66	.55, .78	.51, .81	38.22	7.85*
Chair sit-and-reach	Low	2	2	134	.33	.35	.14, .56	-.13, .83	16.42	12.18*
	High	1	1	48	.81	.86	.79, .94	.86, .86	100.00	.00

Note. *K*, number of studies; *n*, number of *r*s; *N*, total sample size; r_c, overall weighted mean of *r* corrected for sampling error only; r_p, overall weighted mean of *r* corrected for sampling error and measurement error; [a] 95% confidence interval; [b] 95% credibly interval [c] Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of sit-and-reach tests. † Because some studies mixed genders or hamstring extensibility levels were missed, the overall *n* for these categories is lower for some sit-and-reach tests. * $p < 0.05$

the three criteria was met in the SR test protocols (except for the Unilateral SR and Chair SR for males, because logically these had only two and one *r* values, respectively), indicating that the criterion-related validity of these SR tests separately for sex was still heterogeneous. Finally, because some studies grouped males and females together, in Table 3 overall *n* of the sex of participants is lower for some SR test protocols.

Age of participants: The results showed that all SR test protocols had a low-to-moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for children (r_p range = 0.32-0.67) and moderate for adults (r_p range = 0.49-0.68) in which all 95% CI did not include the value zero. Of all the examined SR test protocols, only studies of the Classic SR, Modified SR, and Back-saver SR tests were found for both children and adults. The results of the present meta-analysis showed that there was a trend in the mean correlation coefficient reported to be greater for adults than for children in the Classic SR and Modified SR, but not in the Back-saver SR test where the *r* average values were equal. However, in any case, all 95% CI of mean correlation coefficients were overlapped. Furthermore, we should also be cautious because the low numbers of *r* values over the analyses were supported. Finally, according to moderator analysis criteria, at least one of the three criteria was met in most SR test protocols (except for the Modified SR for children, because logically these had only two *r* values), indicating that the criterion-related validity of these SR tests separately for age were still heterogeneous.

Level of hamstring extensibility: The results showed that all SR test protocols had a low-to-moderate mean correlation coefficient of criterion-related validity for participants with low level of hamstring extensibility (< 80º in the average score of the straight leg raise test) (r_p range = 0.35-0.63) and moderate-to-high for participants with a high level of hamstring extensibility (≥ 80º in the average score of the straight leg raise test) (r_p range = 0.58-0.86) in which all 95% CI did not include the value zero. For all examined SR test protocols, there was a trend of the mean correlation coefficient to being greater for participants with high levels of hamstring extensibility than for those with low levels.

However, we have to be aware that, except for the Chair SR test, all the 95% CI of mean correlation coefficients were overlapped, as well as the low numbers of *r* values over the analyses were supported. Additionally, according to moderator analysis criteria, at least one of the three criteria was met in all SR test protocols (except for the Unilateral SR for low levels and Chair SR for high levels because logically these had only one *r* value), indicating that the criterion-related validity of these SR tests separately for level of hamstring extensibility were still heterogeneous. Finally, because several studies failed to identify the level of hamstring extensibility or were ambiguous (i.e., hamstring extensibility scores around 80º shown graphically), in Table 3 overall *n* of level of hamstring extensibility is lower for some SR tests.

Discussion

From its conception, the Classic SR test has been subjected to numerous modifications, often with the aim of improving its validity. However, according to the results of the present meta-analysis, and although we are aware that all the 95% CI of mean correlation coefficients were overlapped, the Classic SR test showed a greater average criterion-related validity coefficient. Hence, if our purpose is to assess hamstring extensibility, it seems that the use of a modification of the classic protocol is not justified.

Specifically, it has been suggested for several years that the Classic SR test did not consider limb length differences (Hoeger et al., 1990). To solve this methodological "problem", Hoeger et al. (1990) proposed the Modified SR, which incorporates a finger-to-box distance to account for proportional differences between legs and arms. In this line, these authors found that adolescents with longer legs relative to arms had poorer performance on the Classic SR test, and the Modified SR negated the concern about disproportionate limb length bias by establishing a relative zero point for each person. Unfortunately, this study failed to address the very important issue of criterion-related validity.

The present meta-analysis showed a greater overall mean criterion-related validity for the Classic SR than for the Modified SR. In addition, for other modifications of SR tests that incorporated fingers-to-box distance (i.e., the Modified back-saver SR and Modified V SR), the average criterion-related validity coefficients were higher for the end scores version than for the modified one. In this line, in most primary studies in which the criterion-related validity of end and differences scores of SR tests was studied among the same sample, coefficients values were slightly greater for traditional protocols (e.g., Ayala et al., 2011; Castro-Piñero et al., 2009b; Lemmink et al., 2003; López-Miñarro et al., 2010a; López-Miñarro et al., 2010b).

Regarding the criterion-related validity for estimating lumbar extensibility, in addition to the low correlation coefficient found, we have to be aware that the Pearson´s zero-order correlation coefficient was considered; therefore, because of the common explanation for hamstring and lumbar extensibility, the "real" criterion-related validity values for estimating lumbar extensibility could be even lower.

Finally, in line with the results of the present meta-analysis, previous primary studies carried out with young adults (López-Miñarro and Rodríguez-García, 2010c) and elderly women (López-Miñarro et al., 2011) found that the level of hamstring extensibility influenced the criterion-related validity of the Classic SR and Toe touch tests. However, due to the fact that in the present meta-analysis the *n* was classified based on the average scores of the straight leg raise test, we were aware that several participants with low hamstring extensibility could be classified as high flexibility and vice versa. This fact could reduce drastically the difference reported in the results of the present meta-analysis.

Strengths and limitations

The meta-analysis is a useful tool to assess the scientific evidence, but an understanding of its strengths and limitations is needed for most appropriate use of this method (Flather et al., 1997). Overall, the main strength of a meta-analysis is that it lets us obtain more reliable population estimates of findings than those of the constituent studies. Therefore, the results of a meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies. Likewise, the meta-analysis represents the best up-to-date approach to describe and summarize the scientific findings of a research area (Hunter and Schmidt, 2004). Lastly, meta-analysis methods can advance in an entire discipline by addressing more general questions in the area (Cooper et al., 2009).

Regarding the strengths of the present meta-analysis, we followed several measures to avoid (or at least to reduce) publication bias. A lot of research studies fail to be published at all, while others are published only in abstract form, conference proceeding, or dissertation but not as scientific articles. Furthermore, research studies with favorable results are far more likely to be published than those with inconclusive results. Likewise, identification of relevant studies may also be difficult because of their publication in less accessible journals. Thus, performing a meta-analysis when a proportion of the relevant data is missing can provide misleading results, and publication bias may spuriously support a hypothesis by continuously selecting favorable results and rejecting unfavorable ones (Flather et al., 1997).

Therefore, to avoid availability bias, we conducted a wide literature search. The potential inclusion of all relevant single studies in the present meta-analysis (i.e., published and unpublished or English and non-English language) by extent and careful searching might clearly help reduce the impact of publication bias in the present meta-analysis. Hence, the inclusion of unpublished and non-English language studies in the literature search is an important strength of the present meta-analysis. Multiple publication bias also exists when the same researchers responsible for multiple publications report the same validity coefficients, derived from the same participants under the same experimental conditions. Thus, in the present meta-analysis all studies by the same authors were thoroughly cross-referenced with each other. Since some selected studies had fully or partially duplicated information, these particular correlations values were not analyzed in the meta-analyses. Lastly, several exploratory analyses were also conducted to detect the presence of publication bias.

Finally, the Hunter-Schmidt´s psychometric meta-analysis approach (2004) was conducted in the present study to obtain the population estimates of criterion-related validity of SR tests. Because sample sizes are never infinite and measures are never perfectly reliable, sampling error and measurement error are always present in all real data. The psychometric meta-analysis approach corrects the observed correlations due both to sampling error and measurement error. Thus, this method is probably one of the best approaches to estimate the population correlation coefficients.

On the other hand, there were some limitations that should be considered when examining the results of the present study. The main limitations of the present meta-analysis were related to the small number of criterion-related validity coefficients found. Firstly, estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis. Secondly, a partially hierarchical breakdown had to be used. The main problem in this kind of analysis is that it might produce quite misleading results due to confounding and interaction effects. We are aware that a fully hierarchical moderator analysis approach may be a more appropriate method to resolve this problem. However, more correlations coefficients would be needed for each level of moderators. For these reasons, the results of the present study should be considered with caution, especially for those SR test protocols from which only a few studies were retrieved. Firmer conclusions should await the accumulation of a larger number of studies (Hunter and Schmidt, 2004).

Another limitation of the present meta-analysis is related to the criterion measures used in the included studies. Joint(s) range of motion measured through radiography seems to be the best criterion measurement to assess flexibility (Gajdosik and Bohannon, 1987), but due to several practical reasons such us high cost, necessity of sophisticated instruments, qualified technicians, or time constraints, the use of this method is limited (Castro-Piñero et al., 2009b). On the other hand, goniometers are relatively easy to obtain, valid and highly accurate instruments to measure joint range of motion; therefore, joint(s) range of motion measured through goniometers has been widely considered a valid and suitable criterion measure of hamstring extensibility (e.g., Ayala et al., 2011; Hartman and Looney, 2003; López-Miñarro and Rodríguez-García, 2010c). In this line, all the previous studies found considered the angular tests measured by goniometers as the criterion measures. However, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman and Looney, 2003). Similarly, although none of the previous studies has used radiography as the criterion measure of lumbar extensibility, they administered tests with a demonstrated high reliability and validity (Macrae and Wright, 1969; Williams et al., 1993).

Another area of concern is that moderator analyses showed that there was still a large amount of unexplained variance after controlling for artefacts and predefined moderators. Studies included in a meta-analysis are expected to vary in a number of ways. Thus, beyond the sampling error and other statistical artefacts, differences between studies (e.g., sample, study design, or tests procedure) undoubtedly affect these results. For example, the straight raise leg test can be measured by different kinds of movements (i.e., active or passive), instruments (e.g., radiography, goniometer or inclinometer), number of researchers, number of repetitions, time of rest between repetitions, and criteria of maximum extensibility. Additionally, in the present meta-analysis different criterion measures were used to estimate the lumbar extensibility. This statistical heterogeneity can be quantified, but there is usually uncertainty about how important the differences really are. Thus, quantifying and accounting for differences between component studies in a meta-analysis remains a substantial methodological problem and a continuing source of debate (Flather et al., 1997).

Finally, coding some study features was problematic due to different reasons. The moderator analysis had missing data in sex categories because some authors mixed males with females in their studies. Hamstring extensibility also had missing data because several authors failed to identify it or it was ambiguous. In addition, because in the present meta-analysis the hamstring extensibility was classified based on the average scores, we are aware that several participants with low hamstring extensibility could be classified as high flexibility and vice versa. Lastly, although participant characteristics such as physical activity levels or sports practice were potentially moderating features, coding for them was not possible because most studies did not identify them.

Conclusion

Overall the SR tests have a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility, but they have a low mean criterion-related validity for estimating lumbar extensibility. The Classic SR test shows the greater average criterion-related validity for estimating hamstring extensibility. The results of the present meta-analysis suggest that the end scores of the classic versions of the SR tests (e.g., the Classic SR) are a better indicator of hamstring extensibility than the modifications that incorporate the fingers-to-box distance (e.g., the Modified SR). Regarding the three potential moderators examined (sex of participants, age of participants, and level of hamstring extensibility), generally females, adults, and participants with high levels of hamstring extensibility tended to have greater mean values of criterion-related validity for estimating hamstring extensibility. However, due to the low number of *r* values found, the fact that almost all the 95% CI of mean correlation coefficients were overlapped, and that criterion-related validity of SR tests within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis.

Therefore, when angular tests such as the straight leg raise or knee extension tests cannot be used, the SR tests seem to be a useful alternative to estimate hamstring extensibility; however, to assess lumbar extensibility other widely used tests such as the Macrae & Wright or Single/Double inclinometer methods should be used. Nevertheless, as in the application of any field fitness test, evaluators must be aware that the results of SR tests are simply an estimation and, therefore, not a direct measure of the hamstring extensibility. On the other hand, when there are a higher number of studies accumulated, a large-sized meta-analysis with a fully hierarchical analysis approach should be carried out. Future research should further study the criterion-related validity of SR tests, especially in modifications of SR tests such as the SR with plantar flexion, among populations such as children or athletes, and go deeply into other related aspects such as the level of hamstring extensibility.

Acknowledgments

We thank Anna Szczesniak and Aliisa Hatten for the English revision. The first author is supported by a research grant from the Spanish Ministry of Education (AP2010-5905).

References

Ayala, F., Sainz de Baranda, P., De Ste Croix, M. and Santonja, F. (2011) Criterion-related validity of four clinical tests used to measure hamstring flexibility in professional futsal players. *Physical Therapy in Sport* **12,** 175-181.

Ayala, F., Sainz de Baranda, P., De Ste Croix, M. and Santonja, F. (2012) Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. *Physical Therapy in Sport* **13,** 219-226.

Baker, A.A. (1985) The relative contribution of flexibility of the back and hamstring muscles in the performance of the sit and reach component of the AAHPERD health related fitness test in girls thirteen to fifteen years of age. Doctoral thesis, University of North Texas, United States.

Baltaci, G., Tunay, V., Besler, A. and Gerçeker, S. (2003) Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. *British Journal of Sports Medicine* **37,** 59-61.

Begg, C.B. and Mazumdar, M. (1994) Operating characteristics of a rank order correlation for publication bias. *Biometrics* **50,** 1088-1101.

Benito Peinado, P.J., Díaz Molina, V., Calderón Montero, F.J., Peinado Lozano, A.B., Martín Caro, C., Álvarez Sánchez, M. and Pérez Tejero, J. (2007) Literature review in exercise physiology: Practical recommendations. *Revista Internacional de Ciencias del Deporte* **6,** 1-11.

Biering-Sorensen, F. (1984) Physical measurements as risk indicator for low-back trouble over a one year period. *Spine* **9,** 106-119.

Book, C. B. (1989) *Validation of the sit-and-reach test*. Doctoral thesis, University of Minnesota, United States.

Bozic, P. R., Pazin, N., Berjan, B. B., Planic, N. M. and Cuk, I. D. (2010) Evaluation of the field tests of flexibility of the lower extremity: Reliability, and the concurrent and factorial validity. *Journal of Strength and Conditioning Research* **24,** 2523-2531.

Cardoso, J.R., Azevedo, N.C.T., Cassano, C.S., Kawano, M.M. and Âmbar G. (2007) Intra and interobserver reliability of angular kinematic analysis of the hip joint during the sit-and-reach test to measure hamstring length in university students. *Revista Brasileña de Fisioterapia* **11,** 133-138.

Castro-Piñero, J., Artero, E.G., España-Romero, V., Ortega, F.B., Sjöström, M. and Ruiz, J.R. (2009a) Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* **44,** 934-943.

Castro-Piñero, J., Chillón, P., Ortega, F.B., Montesinos, J.L., Sjöström, M. and Ruiz, J.R. (2009b) Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *International Journal of Sports Medicine* **30,** 658-662.

Chung, P. K. and Yuen, C. K. (1999) Criterion-related validity of sit-and-reach test in university men in Hong Kong. *Perceptual and Motor Skills* **88,** 304-316.

Cohen, J.A. (1992) Power primer. *Psychological Bulletin* **112,** 155-159.

Cooper, H., Hedges, L.V. and Valentine, J.C. (2009) *The handbook of research synthesis and meta-analysis*. 2nd edition. Sage, New York.

Da Silva Díaz, R. and Gómez-Conesa, A. (2008) Shortened hamstring syndrome. *Fisioterapia* **30,** 186-193.

Davis, D.S., Quinn, R.O., Whiteman, C.T., Williams, J.D. and Young, C.R. (2008) Concurrent validity of four clinical tests used to measure hamstring flexibility. *Journal of Strength and Conditioning Research* **22,** 583-588.

Erkula, G., Demirkan, F., Kilic, B.A. and Kiter, E. (2002) Hamstring shortening in healthy adults. *Journal of Back and Musculoskeletal Rehabilitation* **16,** 77-81.

Esola, M.A., McClure, P.W., Fitzgerald, G.K. and Siegler, S. (1996) Analysis of lumbar spine and hip motion during forward bending in subjects with and without a history of low back pain. *Spine* **21,** 71-78.

Fisk, J.W., Baigent, M.L. and Hill, P.D. (1984) Scheuermann´s disease. Clinical and radiological survey of 17 and 18 year olds. *American Journal of Physical Medicine* **63,** 18-30.

Flather, M.D., Farkouh, M.E., Pogue, J.M. and Yusuf, S. (1997) Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials* **18,** 568-579.

Gajdosik, R.L. and Bohannon, R.W. (1987) Clinical measurement of range of motion: Review of goniometry emphasizing reliability and validity. *Physical Therapy* **67,** 1867-1872.

García, S.C. (1995) *Validity of the sit-and-reach test for male and female adolescents*. Doctoral thesis, University of Oregon, United States.

Hartman, J.G. and Looney, M. (2003) Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science* **7,** 71-87.

Harvey, J. and Tanner, S. (1991) Low back pain in young athletes: A practical approach. *Sports Medicine* **12,** 394-406.

Hoeger, W.W., Hopkins, D.R., Button, S. and Palmer, T.A. (1990) Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science* **2,** 156-162.

Holt, L.E., Pelma, T.W. and Burke, D.G. (1999) Modifications to the standard sit-and-reach flexibility protocol. *Journal of Athletic Training* **34,** 43-47.

Hui, S.S. and Yuen, P.Y. (2000) Validity of the modified back-saver sit-and-reach test: A comparison with other protocols. *Medicine & Science in Sports & Exercise* **32,** 1655-1659.

Hui, S.C., Yuen, P.Y., Morrow, J.R. and Jackson, A.W. (1999) Comparison of the criterion-related validity of sit-and-reach tests with and without limb length adjustment in Asian adults. *Research Quarterly for Exercise and Sport* **70,** 401-406.

Hunter, J.E. and Schmidt, F.L. (2004) *Methods of meta-analysis: Correcting error and bias in research findings*. 2nd edition. Sage, Newbury Park.

Jackson, A.W. and Baker, A.A. (1986) The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in young females. *Research Quarterly for Exercise and Sport* **57,** 183-186.

Jackson, A. and Langford, N.J. (1989) The criterion-related validity of the sit and reach test: Replication and extension of previous findings. *Research Quarterly for Exercise and Sport* **60,** 384-387.

Jones, C. J., Rikli, R. E., Max, J. and Noffal, G. (1998) The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adult. *Research Quarterly for Exercise and Sport* **69,** 338-343.

Kanbur, N. O., Düzgün, I., Derman, O. and Baltaci, G. (2005) Do sexual maturation stages affect flexibility in adolescent boys aged 14 years? *The Journal of Sports Medicine and Physical Fitness* **45,** 53-57.

Kawano, M.M., Ambar, G., Oliveira, B.I.R., Boer, M.C., Cardoso, A.P. R.G. and Cardoso, J.R. (2010) Influence of the gastrocnemios muscle on the sit-and-reach test assessed by angular kinematic analysis. *Revista Brasileña de Fisioterapia* **14,** 10-15.

Kendall, F.P., McCreary, E.K., Provance, P.G., Rodgers, M.M. and Romani, W.A. (2005) *Muscles: Testing and function with pos-*

ture and pain. 5th edition. Lippincott, Williams, & Wilkins, Baltimore.
Langford, N.J. (1987) The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in adult males and females. Doctoral thesis, North Texas State University, United States.
Lemmink, K.A.P.M., Kemper, H.C.G., de Greef, M.H.G., Rispens, P. and Stevens, M. (2003) The validity of the sit-and-reach test and the modified sit-and-reach test in middle-aged to older men and women. *Research Quarterly for Exercise and Sport* **74,** 331-336.
Liemohn, W., Martin, S.B. and Pariser, G.L. (1997) The effect of ankle posture on sit-and-reach test performance. *Journal of Strength and Conditioning Research* **11,** 239-241.
Liemohn, W., Sharpe, G. L. and Wasserman, J. F. (1994) Criterion related validity of the sit-and-reach test. *Journal of Strength and Conditioning Research* **8,** 91-94.
Light, R.J. and Pillemer, D.B. (1984) *Summing up: The science of reviewing research*. Harvard University Press, Cambridge.
Lipsey, M.W. and Wilson, D.B. (2001) *Practical meta-analysis*. Sage, Newbury Park.
López Miñarro, P.A., Ferragut Fiol, C., Alacid Cárceles, F., Yuste Lucas, J.L. and García Ibarra, A. (2008a) Validity of sit-and-reach and toe-touch tests in the evaluation of hamstring muscle length in young paddlers. *Apunts* **43,** 24-29.
López Miñarro, P.A., Rodríguez García, P.L., Yuste, J.L., Alacid, F., Ferragut, C. and García Ibarra, A. (2008b) Validity of the lumbo-sacral position in bending as measure of hamstring muscle extensibility on young athletes. *Archivos de Medicina del Deporte* **25,** 103-110.
López Miñarro, P.A., Sainz de Baranda Andújar, P., Yuste Lucas, J.L. and Rodríguez García, P.L. (2008c) Validity of the unilateral sit-and-reach test as measure of hamstring muscle extensibility. Comparison with other protocols. *Cultura, Ciencia y Deporte* **3,** 87-92.
López-Miñarro, P.A. (2010) Criterion-related validity of the lumbo-horizontal angle in flexion as a measure of hamstring muscle extensibility in young adults. *Cultura, Ciencia y Deporte* **5,** 25-31.
López-Miñarro, P.A. and Alacid, F. (2009) Influence of hamstring muscle extensibility on spinal curvatures in young athletes. *Science & Sports* **25,** 188-193.
López-Miñarro, P.A., Alacid, F., Muyor, J.M. and López, F.J. (2010a) Criterion-related validity of the modified sit-and-reach test as hamstring muscle extensibility measure in young adults. *Kronos* **9,** 39-46.
López-Miñarro, P.A., García Ibarra, A. and Rodríguez García, P.L. (2010b) Comparison between sit-and-reach tests for measuring hamstring muscle extensibility. *Apunts* **99,** 56-64.
López-Miñarro, P.A., Muyor, J.M. and Alacid, F. (2011) Validity of sit-and-reach tests as measures of hamstring extensibility in older women. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte* **11,** 564-572.
López-Miñarro, P.A. and Rodríguez-García, P.L. (2010c) Hamstring muscle extensibility influences the criterion-related validity of sit-and-reach toe-touch tests. *Journal of Strength & Conditioning Research* **24,** 1013-1018.
López-Miñarro, P. A., Sainz de Baranda Andújar, P. and Rodríguez-García, P. L. (2009) A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science and Medicine* **8,** 116-122.
López-Miñarro, P. A., Sáinz de Baranda, P., Rodríguez-García, P. L. and Yuste, J. L. (2008) Comparison between sit-and-reach test and V sit-and-reach test in young adults. *Gazzetta Medica Italiana Archivio per le Scienze Mediche* **167,** 135-142.
López-Miñarro, P. A., Vaquero-Cristóbal, R., Muyor, J. M., Alacid, F. and Isorna, M. (2012) Criterion-related validity of the sit-and-reach test as measure of hamstring extensibility in paddlers. *Cultura, Ciencia y Deporte* **7,** 95-101.
Macrae, I.F. and Wright, V. (1969) Measurement of back movement. *Annals of Rheumatic Disease* **28,** 584-589.
McHugh, M.P., Kremenic, I.J., Fox, M. B. and Gleim, G.W. (1998) The role of mechanical and neural restraints to joint range of motion during passive stretch. *Medicine and Science in Sports and Exercise* **30,** 928-932.
Mier, C.M. (2011) Accuracy and feasibility of video analysis for assessing hamstring flexibility and validity of the sit-and-reach test. *Research Quarterly for Exercise and Sport* **82,** 617-623.
Mierau, D., Cassidy, J.D. and Yong-Hing, K. (1989) Low-back pain and straight in children and adolescents. *Spine* **14,** 526-528.
Minkler, S. and Patterson, P. (1994) The validity of the modified sit-and-reach test in college-age students. *Research Quarterly for Exercise and Sport* **65,** 189-192.
Miyazaki, J., Murata, S., Horie, J. and Suzuki, S. (2010) Relationship between the sit-and-reach distance and spinal mobility and straight leg raising range. *Rigakuryoho Kagaku* **25,** 683-686.
Orloff, H. A. (1988) *Standardization, reliability and validation of flexibility field testing protocols*. Doctoral thesis, University of Kansas, United States.
Patterson, P., Wiksten, D. L., Ray, L., Flanders, C. and Sanphy, D. (1996) The validity and reliability of the back saver sit-and-reach test in middle school girls and boys. *Research Quarterly for Exercise and Sport* **67,** 448-451.
Rodríguez-García, P.L., López-Miñarro, P.A., Yuste, J.L. and Sáinz de Baranda, P. (2008) Comparison of hamstring criterion-related validity, sagittal spinal curvatures, pelvic tilt and score between sit-and-reach and toe-touch tests in athletes. *Medicina dello Sport* **61,** 11-20.
Rosenthal, R. (1979) The "file drawer problem" and tolerance for null results. *Psychological Bulletin* **33,** 1005-1008.
Santonja Medina, F., Ferrer López, V. and González-Moro, I.M. (1995) Clinical examination of hasmtring tightness. *Selección* **4,** 81-91.
Simoneau, G. G. (1998) The impact of various anthropometric and flexibility measurements on the sit-and-reach test. *Journal of Strength and Conditioning Research* **12,** 232-237.
Smith, J.F. and Miller, C.V. (1985) The effect of head position on sit-and-reach performance. *Research Quarterly for Exercise and Sport* **56,** 84-85.
Standaert, C. J. and Herring, S. A. (2000) Spondylolysis: A critical review. *British Journal of Sports Medicine* **34,** 415-422.
Williams, R., Binkley, J., Bloch, R., Goldsmith, C.H. and Minuk, T. (1993) Reliability of the Modified-modified Schöber and Double inclinometer methods for measuring lumbar flexion and extension. *Physical Therapy* **73,** 26-37.
Yuen, P. Y., and Hui, S. C. (1998) Are difference scores better predictor of flexibility than end scores in sit-and-reach tests? *Medicine & Science in Sports & Exercise* **30,** S125.

Key points

- Overall sit-and-reach tests have a moderate mean criterion-related validity for estimating hamstring extensibility, but they have a low mean validity for estimating lumbar extensibility.
- Among all the sit-and-reach test protocols, the Classic sit-and-reach test seems to be the best option to estimate hamstring extensibility.
- End scores (e.g., the Classic sit-and-reach test) are a better indicator of hamstring extensibility than the modifications that incorporate fingers-to-box distance (e.g., the Modified sit-and-reach test).
- When angular tests such as straight leg raise or knee extension tests cannot be used, sit-and-reach tests seem to be a useful field test alternative to estimate hamstring extensibility, but not to estimate lumbar extensibility.

AUTHORS BIOGRAPHY

Daniel MAYORGA-VEGA

Employment

Research fellow, Department of Physical Education and Sport, University of Granada, Spain.

Degree

PhD student

Research interests

Measurement and evaluation, health-related physical fitness, physical education-based interventions, motivation toward physical activity.

E-mail: dmayorgavega@gmail.com

Rafael MERINO-MARBAN

Employment

Tenured Professor, Department of Didactics of Musical, Plastic and Corporal Expression, University of Malaga, Spain

Degree

PhD

Research interests

Health-related physical fitness, physical education-based interventions, kinesio tape.

E-mail: rmerino@uma.es

Jesús VICIANA

Employment

Tenured Professor, Department of Physical Education and Sport, University of Granada, Spain.

Degree

PhD

Research interests

Planning in physical education, physical education-based interventions, motivation toward physical activity.

E-mail: jviciana@ugr.es

✉ **Daniel Mayorga-Vega**

Department of Physical Education and Sport, University of Granada, Spain

Paper VIII

CRITERION-RELATED VALIDITY OF TOE-TOUCH TEST FOR ESTIMATING HAMSTRING EXTENSIBILITY: A META-ANALYSIS

Mayorga-Vega, D., Viciana, J., Cocca, A., & Merino-Marban, R.

Journal of Human Sport & Exercise

2014, *9*(1), 188-200.

Original Article

Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis

DANIEL MAYORGA-VEGA[1], JESÚS VICIANA[1], ARMANDO COCCA[1], RAFAEL MERINO-MARBAN[2]

[1] *Department of Physical Education and Sport, Faculty of Sciences of Physical Activity and Sport, University of Granada, Spain*

[2] *Department of Didactics of Musical, Plastic and Body Expression, Faculty of Education Sciences, University of Malaga, Spain*

ABSTRACT

Mayorga-Vega, D., Viciana, J., Cocca, A., & Merino-Marban, R. (2014). Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis *J. Hum. Sport Exerc., 9*(1), pp. 188-200. The main purpose of the present meta-analysis was to examine the scientific literature on the criterion-related validity of the toe-touch test for estimating hamstring extensibility. For this purpose relevant studies were searched from five electronic databases dated up through September 2012. Primary outcomes of criterion-related validity were Pearson´s zero-order correlation coefficients (*r*) between the toe-touch test and hamstring extensibility criterion measure. Then, from the included studies, the Hunter-Schmidt´s psychometric meta-analysis approach was conducted to estimate population criterion-related validity of the toe-touch test. Firstly, the corrected correlation mean (r_p), unaffected by sampling error and measurement error, was calculated. Subsequently, the three potential moderator variables (sex of participants, age of participants, and level of hamstring extensibility) were examined by a partially hierarchical analysis. Of the six studies included in the present meta-analysis, 12 correlations values were retrieved. The overall results showed that the toe-touch test have a moderate mean criterion-related validity for estimating hamstring extensibility (r_p = 0.66, 0.54-0.79). Generally, females, children and individuals with high levels of hamstring extensibility seem to have greater mean values of criterion-related validity for estimating hamstring extensibility. However, due to the low number of *r* values found, the fact that almost all the 95% CIs of mean correlation coefficients were overlapped, and that criterion-related validity of the toe-touch test within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis. When the use of the angular tests is limited, the toe-touch test seems to be a useful alternative to estimate hamstring extensibility. **Key words**: CONCURRENT VALIDITY, RANGE OF MOTION, FLEXIBILITY, STAND AND REACH TEST, LINEAL TEST, FIELD-BASED PHYSICAL FITNESS TEST, RESEARCH SYNTHESIS, SYSTEMATIC REVIEW.

Corresponding author. Facultad de Ciencias de la Actividad Física y el Deporte, Crta. de Alfacar, s/n, 18011, Granada, Spain
E-mail: dmayorgavega@gmail.com
Submitted for publication September 2012
Accepted for publication September 2013
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

doi:10.4100/jhse.2014.91.18

INTRODUCTION

Hamstring extensibility is a well-recognized health-related physical fitness marker that plays an important role in protecting the spine from possible risks and, therefore, allowing people to execute the normal daily living activities and social functioning (Roth-Isigkeit et al., 2005; Sato et al., 2008). For example, reduced hamstring extensibility conditions several spine disorders such as thoracic hyperkyphosis (Fisk et al., 1984), spondylolysis (Standaert & Herring, 2000), disc herniation (Harvey & Tanner, 1991), changes in lumbopelvic rhythm (López-Miñarro & Alacid, 2009) and low back pain (Sjölie, 2004). Additionally, individuals with shortened hamstring muscles present gait limitations, increased risk of falls, and susceptibility to musculoskeletal injuries (Erkula et al., 2002; Jones et al., 1998).

Currently there are different kinds of tests in order to examine the levels of hamstring extensibility. On one hand, the angular tests that specifically measure hip flexion with knee extended (straight leg raise test) (American Academy of Orthopaedic Surgeons, 1996), or the knee extension range with the hip flexed to 90 degrees (knee extension test) (Hartman & Looney, 2003), have been considered as the criterion measures of hamstring extensibility. However, due to the necessity of relatively sophisticated instruments, qualified technicians, and time constraints, the use of these angular tests seem to be limited in several settings (Castro-Piñero et al., 2009b). On the other hand, the lineal tests are characterized by having a simple procedure, being easy to administer, requiring minimal skills training for their application and a very affordable equipment to perform them (Castro-Piñero et al., 2009b; López Miñarro et al., 2008c). Therefore, in contrast with the angular tests, the lineal tests allow the evaluation of a large number of people in a short space of time.

The classic sit-and-reach test (also called traditional or standard sit-and-reach), originally designed by Wells & Dillon (1952), is probably the most widely used lineal test of flexibility in exercise science laboratories, physical education classes, and commercial fitness centres (Cepero et al., 2011; Holt et al., 1999; Mirzaei et al., 2011). Two years later, Kraus & Hirschland (1954) designed the toe-touch (TT) test in which the individuals were assessed standing instead of sitting on the floor like in the classic sit-and-reach test. The most common assumption when interpreting the results of TT test is that individuals with better scores possess a higher level of hamstring extensibility than those with lower scores (Muyor et al., 2012; Sainz de Baranda et al., 2006).

Nevertheless, the primary studies examining the criterion-related validity of the TT test for estimating hamstring extensibility have shown inconclusive results (López-Miñarro et al., 2010c; Rodríguez-García et al., 2008). Each primary study that is published about criterion-related validity of the TT test only constitutes as a single piece of a constantly growing body of evidence (Cooper et al., 2009). To clarify the often conflicting results found in the scientific literature, researchers have to conduct meta-analyses (Cooper et al., 2009; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). Therefore, the meta-analyses remain a useful tool for the evaluation of evidence (Flather et al., 1997).

In this line, recently Mayorga-Vega et al. (2014) carried out a meta-analysis about the criterion-related validity of the sit-and-reach tests. Beyond the simple but important function of describing and summarizing the scientific findings of this research area, the main contribution of the above mentioned meta-analysis was to estimate as accurately as possible the population parameters. Therefore, the results of the Mayorga-Vega´s et al. (2014) meta-analysis let us generalize the research findings, as well as test hypotheses that may have never been tested in primary studies.

Unfortunately, to our knowledge there are no meta-analyses addressing the criterion-related validity of the TT test. Consequently, the main purpose of the present study was to examine the scientific literature on criterion-related validity of the TT test for estimating hamstring extensibility. More specifically, the objectives of this meta-analysis were: (a) to describe and summarize the up-to-date scientific findings of criterion-related validity of the TT test for estimating hamstring extensibility; (b) to estimate the overall population mean of the criterion-related validity coefficients of the TT test for estimating hamstring extensibility; and (c) to examine the influence of some study features (sex of the participants, age of participants, and level of hamstring extensibility) in criterion-related validity coefficients of the TT test.

MATERIAL AND METHODS

All the methodological procedure followed in the present study was based on the Mayorga-Vega´s et al. (2014) meta-analysis. See the methods section of the mentioned manuscript for more detailed information.

Search strategy

The following five electronic databases were searched from their inception through September 2012: SportDiscus, Scopus, Medline, Pubmed, and Web of Science. The search terms used were based on expressions related to the TT test (toe touch, stand and reach, finger to floor distance, fingertip floor distance). The terms of TT test were combined together with the Boolean operator "OR". Since the keywords consisted of more than one word they were enclosed in quotes. In addition, the reference lists of all included papers were manually searched.

Selection criteria

The selection criteria to identify studies that examined the criterion-related validity of the TT test for estimating hamstring extensibility were: (a) studies with apparently healthy participants who did not present any injury, physical and/or mental disabilities; (b) studies with the TT test that yielded the values of the maximum reach of the fingertips, and (c) studies in which hamstring extensibility criterion measurements used are widely accepted in the scientific literature (i.e., the angular tests straight leg raise or knee extension). In addition to papers, master/doctoral dissertations and conference proceedings were also accepted. No language or publication date restrictions were imposed.

Coding study features

For this meta-analysis, data were collected from studies that reported relationships between the TT test and a hamstring extensibility criterion measure with apparently healthy participants of any age. From each selected study the following data were coded: Study identity number, sample size (*n*), sex of participants (1 = males; 2 = females), age of participants (1 = children, < 18 years; 2 = adults, ≥ 18 years), criterion-related validity result (Pearson´s *r* correlation coefficient value), reliability of the TT test (intraclass correlation coefficient value), reliability of the hamstring extensibility criterion measure (intraclass correlation coefficient value), and the level of hamstring extensibility (1 = low level of hamstring extensibility, < 80º in the straight leg raise test; 2 = high level of hamstring extensibility, < 80º in the straight leg raise test) (Kendall et al., 2005).

Since identification of the study features is usually explicitly stated in each of the primary articles, use of more than one rater was deemed unnecessary. For a study to be included in this meta-analysis, sample size, a widely accepted hamstring extensibility criterion measure and Pearson's *r* value were considered to be critical. In the event that the authors mixed subgroups of a study feature (e.g., males mixed with females) or failed to identify a study feature (e.g., reliability scores) the data was omitted. When in the same study data were expressed for both legs separately, the average value of the coefficients was coded.

Data analyses

In the present study, Pearson´s zero-order correlation coefficient (r) was considered the unit of measure as an indication of criterion-related validity of the TT test, which represents the strength of association between the estimate of TT test and the hamstring extensibility criterion measure. If a single study reported more than on r value from different subsamples (e.g., males and females), we assumed each r value form different subsamples to be independent from each other and included them in a single meta-analysis (Lipsey & Wilson, 2001).

Publication bias: In addition to the followed search strategy and selection criteria to avoid availability bias, an examination of the selected studies was carried out to avoid a potential duplication of information retrieved. Since some selected studies had full duplicated information, these particular r correlations values were not analyzed in the meta-analyses. Furthermore, before computing correlations, several exploratory analyses were also conducted to detect the presence of publication bias.

Firstly, a file drawer analysis based on effect size was performed to estimate the number of unlocated studies averaging null results (r = 0) that would have to exist to bring the mean effect size (r_p) down to the small mean r value (Rosenthal, 1979). According to Cohen´s guidelines (1992), the correlation coefficient was interpreted as small when $r < 0.30$. Secondly, according to Light and Pillemer´s graphic method (1984), the scatter plots of correlations coefficients against sample size for the TT test related to hamstring extensibility were analyzed. Finally, with the objective of quantifying the outcomes of the scatter plots, as suggested by Begg & Mazumdar (1994), a Spearman´s rank order correlation between r values and sample size was calculated.

Computation of correlations: The Hunter-Schmidt´s psychometric meta-analysis approach was conducted to obtain the population estimates of the criterion-related validity of the TT test (Hunter & Schmidt, 2004). This approach estimates the population correlation by individually correcting the observed correlations due to various artefacts such as sampling error and measurement error. First, the “bare-bone” mean r (r_c), corrected for only sampling error, was calculated by weighting each r with the respective sample size when aggregating them into r_c. Then, we calculated the corrected mean r at the population level (r_p) that was unaffected by both sampling error and measurement error. The resulting mean correlation corrected for sampling error and measurement error is offered as the best estimate of the population parameter.

In order to correct the measurement errors, the reliability coefficients (intraclass correlation coefficients) of the TT and criterion measure tests were used. Because the reliability coefficients were not available for all of the included studies, the unknown reliability values were previously estimated for each test. The median of the all reported reliability coefficients for the TT and criterion measure tests was used. Finally, the 95% confidence intervals of r_p (95% CI) were calculated.

Moderator analysis: In the present meta-analysis, due to the low number of r values found, partially hierarchical analyses of moderator variables were carried out. According to Hunter and Schmidt (2004), to determine the presence of moderator effects which may affect overall criterion-related validity of the TT test (r_p), three different criteria were simultaneously examined: (a) the 95% credibly interval (95% CV) is relatively large or includes the value zero; (b) the percentage of variance accounted for by statistical artefacts is less than 75% of the observed variance in r_p; and (c) the *Q* homogeneity statistic is statistically significant ($p < 0.05$).

If at least one of the three criteria were met, we concluded that the results could be affected by moderator effects. In case of the presence of moderator effects, criterion-related validity values of the TT test were analyzed separately by: (a) sex of participants (i.e., male and female); (b) age of participants (i.e., children and adults); and (c) level of hamstring extensibility (i.e., low average level and high average level).

RESULTS AND DISCUSSION

Study description

Figure 1 shows a flow chart of the study selection process. Of the 473 literature search results, 17 potentially relevant publications were identified and retrieved for a more detailed evaluation. Finally, due to duplication issues, of the eight studies that met the inclusion criteria, only six studies were included in the present meta-analysis.

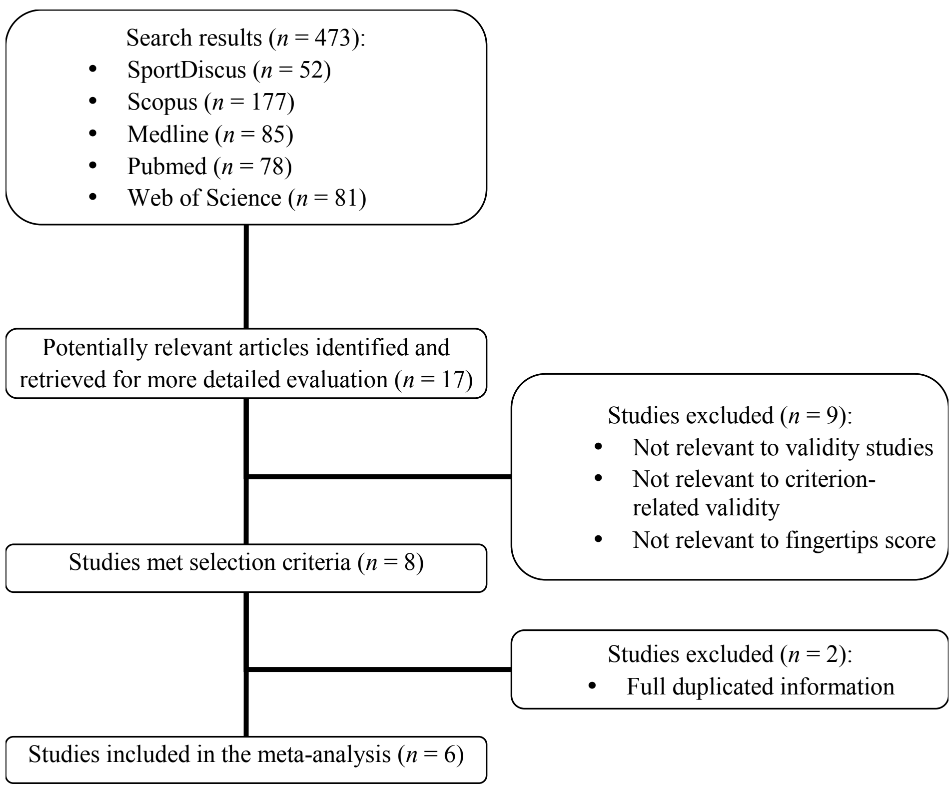


Figure 1. *Flow chart of studies selection process*

Table 1 presents a summary of the retrieved studies of criterion-related validity of the TT test for estimating hamstring extensibility. A total of 15 criterion-related validity coefficients (*r*) for the TT test were retrieved, ranging from low ($r = 0.25$) to high values ($r = 0.92$). A total sample of 1,307 participants (794 males and 513 females) was retrieved. The average age of participants ranged from 13.3 (± 0.6) to 65.3 (± 9.1) years old. Six studies examined the adults and two the children.

Table 1. *Summary of studies of criterion-related validity of the toe-touch test for estimating hamstring extensibility*

Reference	Sample	*n*	Age (years)	Criterion measure	Criterion-related validity (*r*)	
					♂	♀
Ayala et al. (2011)	Professional futsal players	♂=55 ♀=48	26.0 ± 4.5 23.0 ± 5.3	PSLR	0.25	0.92*
Ayala et al. (2012)	Recreationally active university students	♂=156 ♀=87	21.3 ± 2.5 20.7 ± 1.6	PSLR	0.70*	
López Miñarro et al. (2008a)	Canoeists	♂=44 ♀=22	13.3 ± 0.6	PSLR	0.73*-0.66*	0.78*-0.85*
López Miñarro et al. (2008b)	Canoeists	?=66	13.3 ± 0.6	PSLR	**0.73*-0.73***	
López-Miñarro et al. (2010b)	University students	♂=130 ♀=110	22.9 ± 3.2 23.2 ± 4.5	PSLR	**0.57*-0.62***	**0.72*-0.75***
López-Miñarro et al. (2011)	Older women: Low, moderate and high flexibility	♀=36	65.3 ± 9.1	PSLR		0.48*-0.46*
		♀=35				0.61*-0.59*
		♀=35				0.78*-0.76*
López-Miñarro and Rodríguez-García (2010c)	Recreationally active university students: Low and normal flexibility	♂=120	22.9 ± 3.6	PSLR	0.28*-0.40*	
		♂=120			0.60*-0.55*	
Rodríguez-García et al. (2008)	Fit sports activities practitioners	♂=125 ♀=118	22.9 ± 3.2 23.2 ± 4.5	PSLR	0.57*-0.62*	0.72*-0.75*

Note. This table includes all the studies that met selection criteria, however, full information of two studies was not included in the meta-analysis due to duplication issues (in bold); ♂, males; ♀, females; ?, information unavailable; PSLR, Passive straight leg raise test; 0.xx-0.xx, Pearson´s *r* for the left and right leg, respectively.
* Pearson´s *r* statistically significant at $p < 0.05$

Publication bias

Due to two studies having fully duplicated information, these *r* coefficients values were not analyzed in the present meta-analyses despite the fact that these studies met the selection criteria. López Miñarro´s et al. (2008b) study information (males mixed with females) were not computed because the same data were also published with males and females separately (López Miñarro et al., 2008a). Additionally, López-Miñarro´s et al. (2010b) study information (for both males and females) was not computed because the same data had been published previously in other journal (although the sample size was slightly different, it was considered a duplication issue because all the other information was equal) (Rodríguez-García et al., 2008). Pearson´s *r* correlation values of selected studies that were excluded for the meta-analysis are indicated (in bold) in Table 1.

Subsequently, several exploratory analyses were conducted to detect the presence of publication bias. Firstly, the file drawer analysis based on the effect size was calculated. The results of the file drawer analysis showed that 14 unlocated studies averaging null results (r = 0) would have to exist to bring the mean r_p down to 0.29. This is a large number of “lost” studies, especially if we aware of the percentage of unlocated/located studies (117%). Hence, we concluded that it was unlikely that there would be this particular number of “lost” studies.

Then, the Figure 2 shows the scatter plot of sample size against criterion-related validity coefficients of the TT test for estimating hamstring extensibility. Based on the statistical significance of the studies, if there would be publication bias in the scatter plot the small-sample studies reporting small r values would be disproportionately absent because they are studies that would fail to attain statistical significance. Therefore, according to this graphic method, the figure suggested that in the present study there was an absence of publication bias. Finally, in line with the graphic method, the results of Spearman´s rank order correlation between r values and sample size did not show a statistically significant correlation for estimating hamstring extensibility (r = - 0.35, p = 0.263). Similarly, in the presence of publication bias, this correlation should be statistically significant negative due to the absence of small-sample studies in the lower left hand corner.

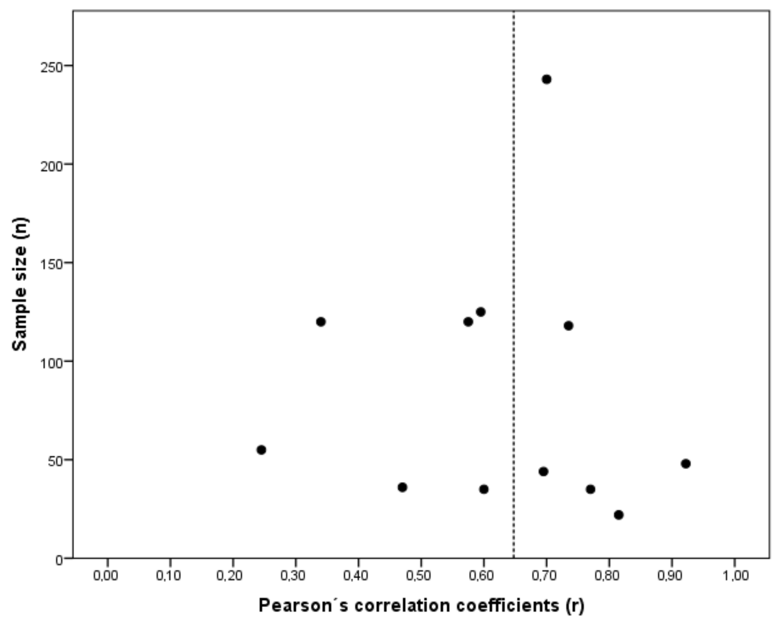


Figure 2. *Scatter plot of sample size and criterion-related validity coefficients (r) of the toe-touch test for estimating hamstring extensibility. Dashed line represents the median value of the validity coefficients.*

Criterion-related validity

Table 2 reports the number of studies (K), the cumulative number of r values (n), the total sample size accumulated (N), the weighted mean of r corrected for sampling error only (r_c), the weighted mean of r corrected for both sampling error and measurement error (r_p), as well as the 95% CI of the criterion-related validity correlation coefficients (r_p) for the TT test for estimating hamstring extensibility. In addition, to detect the presence of moderator effects which may affect the criterion-related validity of the TT test, the 95% CV, the percentage of variance accounted for by statistical artefacts, and the Q homogeneity statistic were calculated.

Table 2. *Results of the meta-analysis for the criterion-related validity correlation coefficients of the toe-touch test for estimating hamstring extensibility*

	K	n	N	r_c	r_p	95% CI[a]	95% CV[b]	% of variance[c]	Q statistic
Overall	6	12	1,001	0.61	0.66	0.54-0.79	0.32-1.00	10.96	109.50*
Sex of participants[d]									
Males	4	5	464	0.49	0.52	0.37-0.67	0.25-0.79	23.49	21.29*
Females	4	6	294	0.73	0.77	0.66-0.89	0.53-1.00	17.68	33.94*
Age of participants									
Children	1	2	66	0.74	0.78	0.65-0.92	0.78-0.78	100.00	1.51
Adults	5	10	935	0.60	0.66	0.54-0.77	0.30-1.00	9.82	101.88*
Level of hamstring extensibility									
Low	5	6	428	0.52	0.55	0.38-0.71	0.15-0.94	14.79	40.57*
High	5	6	573	0.68	0.75	0.66-0.84	0.61-0.89	28.13	21.33*

Note. K, number of studies; n, number of rs; N, total sample size; r_c, overall weighted mean of r corrected for sampling error only; r_p, overall weighted mean of r corrected for sampling error and measurement error; [a] 95% confidence interval; [b] 95% credibly interval; [c] Percentage of variance accounted for by statistical artefacts including sampling error and measurement error of toe-touch test. [d] Because a study mixed males and females, the overall n for this category is lower. * $p < 0.05$

The overall result showed that the TT test had a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility (r_p = 0.66, 0.54-0.79) in which the 95% CI did not include the value zero. However, we should be extremely cautious because the low numbers of r values over the present meta-analysis were supported. In line with the present meta-analysis, recently Mayorga-Vega et al. (2014) carried out a meta-analysis about the criterion-related validity of the sit-and-reach tests for estimating hamstring and lumbar extensibility. These authors found that overall the classic sit-and-reach tests also had a moderate criterion-related validity for estimating hamstring extensibility (r_p = 0.67, 0.55-0.80). Additionally, because the percentage of variance accounted for by statistical artefacts was less than 75%, the Q homogeneity statistic was statistically significant (p < 0.05), and the 95% CV was large, we concluded that the results could be affected by moderator effects. Therefore, follow-up moderator analyses were conducted using predefined moderators as it was hypothesized in the present study.

The classic sit-and-reach test is probably the most widely used lineal measure of flexibility in exercise science laboratories, physical education classes, and commercial fitness centres (Cepero et al., 2011; Holt et al., 1999; Mirzaei et al., 2011). In this line, Castro-Piñero et al. (2009a) carried out a systematic review of the criterion-related validity of field-based fitness battery tests worldwide for youth. These authors found that 91% (10 of 11) of battery tests that included the flexibility assessment proposed the classic sit-and-reach test. However, in none of these batteries the TT test was proposed. However, according to the results of the present meta-analysis, if the purpose is to assess hamstring extensibility, it seems that the use of one test over the other is not justified.

Moderator analyses
Table 2 reports the results of moderator analyses to examine the effects of the sex of the participants (i.e., male and female), the age of the participants (i.e., children and adults), and the level of hamstring extensibility (i.e., low average level, < 80º, and high average level, ≥ 80º) on overall criterion-related validity correlation coefficient of the TT test for estimating hamstring extensibility.

Sex of participants: The results of the present study showed that the TT test had a moderate-low mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for males (r_p = 0.52, 0.37-0.67) and moderate-high for females (r_p = 0.77, 0.66-0.89) in which the 95% CIs did not include the value zero. Additionally, aside from this considerably greater value for females compared to males, the 95% CIs of mean correlation coefficients were hardly overlapped. However, according to moderator analyses criteria, all the criteria were met in both males and females, indicating that the criterion-related validity of the TT test separately for sex was still heterogeneous. Finally, because a study grouped males and females together, in Table 3 the overall *n* for the sex of participants is lower.

The results of the present meta-analysis suggest that the sex of participants affects the criterion-related validity of the TT test for estimating hamstring extensibility. Therefore, it seems that the use of the TT test is more appropriate among females than among males. In this line, most studies either examined only one sex or both but without grouping them probably because they intuited that the sex of participants was a feature than would affect the relationship between the TT test and criterion measure (Ayala et al., 2011; López Miñarro et al., 2008a; López-Miñarro et al., 2011; López-Miñarro et al., 2010a; López-Miñarro et al., 2010b; Rodríguez-García et al., 2008). Similarly, Mayorga-Vega et al. (2014) found that the classic sit-and-reach showed a trend to be more valid for females (r_p = 0.70, 0.58-0.82) than for males (r_p = 0.64, 0.50-0.78). However, in contrast with the present meta-analysis, in the Mayorga-Vega´s et al. (2014) study these differences were really small and the 95% CIs were overlapped.

Age of participants: The results of the present meta-analysis showed that the TT test had a moderate-high mean correlation coefficient of criterion-related validity for estimating hamstring extensibility for children (r_p = 0.78, 0.65-0.92) and moderate for adults (r_p = 0.66, 0.54-0.77) in which the 95% CIs did not include the value zero. These results suggested that the criterion-related validity of the TT test is greater for children than for adults. However, the 95% CIs of mean correlation coefficients were overlapped. Furthermore, we should be extremely cautious because the analyses for children were supported only over two *r* values. In this line, future studies with children about the criterion-related validity of the TT test are required. In this line, Mayorga-Vega et al. (2014) did not find out differences in the results of the classic sit-and-reach between age categories. Finally, according to moderator analyses criteria, the three criteria were met among adults (for children clearly were not met because these analyses had only two *r* values), indicating that the criterion-related validity of the TT test separately for age were still heterogeneous.

Level of hamstring extensibility: The results of this study showed that the TT test had a moderate-low mean correlation coefficient of criterion-related validity for participants with low level of hamstring extensibility (< 80º in the average score of the straight leg raise test) (r_p = 0.55, 0.38-0.71) and moderate-high for participants with a high level of hamstring extensibility (≥ 80º in the average score of the straight leg raise test) (r_p = 0.75, 0.66-0.84) in which the 95% CIs did not include the value zero. The results of the present meta-analysis suggested that the criterion-related validity of the TT test is larger for participants with high level of hamstring extensibility than those with low hamstring extensibility. However, we have to be aware that the 95% CIs of mean correlation coefficients were slightly overlapped. Additionally, according to moderator analyses criteria, at least two of the three criteria were met in both categories, indicating that the criterion-related validity of the TT test separately for level of hamstring extensibility were still heterogeneous.

Similarly, in line with the results of the present meta-analysis, previous primary studies carried out with young adults (López-Miñarro & Rodríguez-García, 2010b) and elderly women (López-Miñarro et al., 2011) found that the level of hamstring extensibility influenced the criterion-related validity of the TT test. However, due to the fact that in the present meta-analysis the *n* was classified based on the average scores of the straight leg raise test, we were aware that several participants with low hamstring extensibility could be classified as high flexibility and vice versa. This fact could reduce drastically the difference reported in the results of the present meta-analysis. Although Mayorga-Vega et al. (2014) found the same trend in their meta-analysis with the classic sit-and-reach test, the 95% CIs were overlapped and they also pointed out this methodological limitation.

STRENGTHS AND LIMITATIONS

An extensive revision of the general strengths and limitations of the meta-analyses, as well as specifically in the meta-analysis of the criterion-related validity of the field-based flexibility tests, has been previously published (Mayorga-Vega et al., 2014). Briefly, regarding the strengths of the present meta-analysis, we followed several measures to avoid (or at least to reduce) publication bias. Firstly, to avoid availability bias, we conducted a wide literature search thought several databases without limiting any kind of manuscript (i.e., articles, master/doctoral dissertations, and conference proceedings), language (i.e., English and non-English language) or publication date. Secondly, in the present meta-analysis all the studies by the same authors were thoroughly cross-referenced with each other in order to avoid duplicated information. Lastly, several exploratory analyses were also conducted to detect the potential presence of publication bias.

Another strength of the present meta-analysis is related to the statistical approach used. In the present study, the Hunter-Schmidt´s psychometric meta-analysis approach (2004) was conducted in order to obtain the population estimates of criterion-related validity of the TT test. Since this method estimates the population correlation by individually correcting the observed correlations due to various artefacts such as sampling error and measurement error, it has been considered one of the best meta-analyses approaches.

On the other hand, there were some limitations that should be considered when examining the results of the present meta-analysis. The main limitations were related to the small number of criterion-related validity coefficients found. Firstly, estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis. Secondly, because a partially hierarchical breakdown had to be used, quite misleading results due to confounding and interaction effects might be produced. Therefore, the results of the present study should be considered with caution.

Another limitation of the present meta-analysis is related to the criterion measures used in the included studies. Although all the previous studies found considered the angular tests measured by goniometers as the criterion measures, nowadays some studies have suggested that the criterion measures of hamstring extensibility must be reexamined and readjusted (Cardoso et al., 2007; Hartman & Looney, 2003). Finally, coding some study features was problematic due to different reasons. For instance, because in the present meta-analysis the level of hamstring extensibility was classified based on the average scores, we are aware that several individuals with low hamstring extensibility could be classified as high flexibility and vice versa. Additionally, although participant characteristics such as physical activity levels or sports practice were potentially moderating features, coding for them was not possible because most studies did not identify them.

CONCLUSIONS

Overall the TT test has a moderate mean correlation coefficient of criterion-related validity for estimating hamstring extensibility. The results of the present meta-analysis show that the TT test has the same population estimated criterion-related validity for estimating hamstring extensibility than the widely used lineal test classic sit-and-reach. Regarding the three potential moderators examined (sex of participants, age of participants, and level of hamstring extensibility), generally females, children, and individuals with high levels of hamstring extensibility seem to have greater mean values of criterion-related validity for estimating hamstring extensibility. However, due to the low number of *r* values found, the fact that almost all the 95% CIs of mean correlation coefficients were overlapped, and that criterion-related validity of the TT test within each category was still heterogeneous, we should be cautious with the results of the present meta-analysis.

When angular tests such as the straight leg raise or knee extension tests cannot be used, the TT test seems to be a useful alternative to estimate hamstring extensibility. Nevertheless, as in the application of any field-based fitness test, evaluators must be aware that the results of the TT test is simply an estimation and, therefore, not a direct measure of the hamstring extensibility. On the other hand, when there are a higher number of studies accumulated, a large-sized meta-analysis with a fully hierarchical analysis approach should be carried out. Future research should further study the criterion-related validity of the TT test for estimating hamstring extensibility, especially among populations such as children, and go deeply into other related aspects such as the influence of the level of hamstring extensibility.

ACKNOWLEDGEMENTS

We thank Anna Szczesniak for the English revision. The author Daniel Mayorga-Vega is supported by a research grant from the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

REFERENCES

1. American Academy of Orthopaedic Surgeons (1996). *Joint motion: Method of measuring and recording*. Livingstone: Edinburgh.
2. Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & Santonja, F. (2011). Criterion-related validity of four clinical tests used to measure hamstring flexibility in professional futsal players. *Phys Ther Sport, 12*(4), pp.175-181.
3. Ayala, F., Sainz de Baranda, P., De Ste Croix, M., & Santonja, F. (2012). Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. *Phys Ther Sport*, *13*(4), pp.219-226.
4. Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank order correlation for publication bias. *Biometrics*, *50*(4), pp.1088-1101.
5. Cardoso, J. R., Azevedo, N. C. T., Cassano, C. S., Kawano, M. M., & Âmbar, G. (2007). Intra and interobserver reliability of angular kinematic analysis of the hip joint during the sit-and-reach test to measure hamstring length in university students. *Rev Bras Fisioterapia*, *11*(2), pp.133-138.

6. Castro-Piñero, J., Artero, E. G., España-Romero, V., Ortega, F. B., Sjöström, M., & Ruiz, J. R. (2009a). Criterion-related validity of field-based fitness tests in youth: A systematic review. *Brit J Sport Med*, *44*(13), pp.934-943.
7. Castro-Piñero, J., Chillón, P., Ortega, F. B., Montesinos, J. L., Sjöström, M., & Ruiz, J. R. (2009b). Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *Int J Sport Med*, *30*(9), pp.658-662.
8. Cepero, M., López, R., Suárez-Llorca, C., Andreu-Cabrera, E., & Rojas, F. J. (2011). Fitness test profiles in children aged 8-12 years old in Granada (Spain). *J Hum Sport Exerc, 6*(1), pp.135-145.
9. Cohen, J. A. (1992). Power primer. *Psychological Bulletin*, *112*(1), pp.155-159.
10. Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Sage.
11. Erkula, G., Demirkan, F., Kilic, B. A., & Kiter, E. (2002). Hamstring shortening in healthy adults. *J Back Musculoskelet*, *16*(2), pp.77-81.
12. Fisk, J. W., Baigent, M. L., & Hill, P. D. (1984). Scheuermann´s disease. Clinical and radiological survey of 17 and 18 year olds. *Am J Phys Med*, *63*(1), pp.18-30.
13. Flather, M. D., Farkouh, M. E., Pogue, J. M., & Yusuf, S. (1997). Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Control Clin Trials*, *18*(6), pp.568-579.
14. Hartman, J. G., & Looney, M. (2003). Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Meas Phys Educ Exer Sci*, *7*(2), pp.71-87.
15. Harvey, J., & Tanner, S. (1991). Low back pain in young athletes: A practical approach. *Sports Med*, *12*(6), pp.394-406.
16. Holt, L. E., Pelma, T. W., & Burke, D. G. (1999). Modifications to the standard sit-and-reach flexibility protocol. *J Athl Training*, *34*(1), pp.43-47.
17. Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park: Sage.
18. Jones, C. J., Rikli, R. E., Max, J., & Noffal, G. (1998). The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adult. *Res Q Exercise Sport*, *69*(4), pp.338-343.
19. Kendall, F. P., McCreary, E. K., Provance, P. G., Rodgers, M. M., & Romani, W. A. (2005). *Muscles: Testing and function with posture and pain* (5th ed.). Baltimore: Lippincott, Williams, & Wilkins.
20. Kraus, H., & Hirschland, R. P. (1954). Minimum muscular fitness tests in school children. *The Res Quarterly, 25*(2), pp.178-188.
21. Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge: Harvard University Press.
22. Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Newbury Park: Sage.
23. López Miñarro, P. A., Ferragut Fiol, C., Alacid Cárceles, F., Yuste Lucas, J. L., & García Ibarra, A. (2008a). Validity of sit-and-reach and toe-touch tests in the evaluation of hamstring muscle length in young paddlers. *Apunts*, *43*(157), pp.24-29.
24. López Miñarro, P. A., Rodríguez García, P. L., Yuste, J. L., Alacid, F., Ferragut, C., & García Ibarra, A. (2008b). Validity of the lumbo-sacral position in bending as measure of hamstring muscle extensibility on young athletes. *Arch Med Depor, 25*(124), pp.103-110.
25. López Miñarro, P. A., Sainz de Baranda Andújar, P., Yuste Lucas, J. L., & Rodríguez García, P. L. (2008c). Validity of the unilateral sit-and-reach test as measure of hamstring muscle extensibility. Comparison with other protocols. *Cultura, Ciencia y Deporte*, *3*(8), pp.87-92.
26. López-Miñarro, P. A., & Alacid, F. (2009). Influence of hamstring muscle extensibility on spinal curvatures in young athletes. *Sci Sport*, *25*, pp.188-193.

27. López-Miñarro, P. A., Alacid, F., Muyor, J. M., & López, F. J. (2010a). Criterion-related validity of the modified sit-and-reach test as hamstring muscle extensibility measure in young adults. *Kronos*, *9*(17), pp.39-46.
28. López-Miñarro, P. A., García Ibarra, A., & Rodríguez García, P. L. (2010b). Comparison between sit-and-reach tests for measuring hamstring muscle extensibility. *Apunts*, *99*(1), pp.56-64.
29. López-Miñarro, P.A., Muyor, J. M., & Alacid, F. (2011). Validity of sit-and-reach tests as measures of hamstring extensibility in older women. *Rev Int Med Ciencias Act Fís Deport, 11*(43), pp.564-572.
30. López-Miñarro, P. A., & Rodríguez-García, P. L. (2010c). Hamstring muscle extensibility influences the criterion-related validity of sit-and-reach and toe-touch tests. *J Strength Cond Res*, *24*(4), pp.1013-1018.
31. Mayorga-Vega, D., Merino-Marban, R., & Viciana, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *J Sport Sci Med*. *J Sport Sci Med*, *13*(1), pp.1-14.
32. Mirzaei, B., Curby, D. C., Barbas, I., & Lotfi, N. (2011). Anthropometric and physical fitness traits of four-time World Greco-Roman wrestling champion in relation to national norms: A case study. *J Hum Sport Exerc, 6*(2), pp.406-413.
33. Muyor, J. M., López-Miñarro, P. A. & Casimiro, A.J. (2012). Effect of stretching program in an industrial workplace on hamstring flexibility and sagittal spinal posture of adult women workers: A randomized controlled trial. *J Back Musculoskelet*, *25*(3), pp.161-169.
34. Rodríguez-García, P. L., López-Miñarro, P. A., Yuste, J. L., & Sáinz de Baranda, P. (2008). Comparison of hamstring criterion-related validity, sagittal spinal curvatures, pelvic tilt and score between sit-and-reach and toe-touch tests in athletes. *Med Sport*, *61*(1), pp.11-20.
35. Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychol Bull*, *86*(3), pp.683-641.
36. Roth-Isigkeit, A., Thyen U., Stöven, H., Schwarzenberger, J., & Schmucker, P. (2005). Pain among children and adolescents: Restrictions in daily living and triggering factors. *Pediatrics*, *115*(2), pp.e152-162.
37. Sainz de Baranda, P., Rodríguez, P. L., Santonja, F. M., López, P. A., Andújar, P., Ferrer, V., & Pastor, A. (2006). Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *J Hum Mov Stud, 51*(4), pp.277-289.
38. Sato, T., Ito, T., Hirano, T., Morita, O., Kikuchi, R., Endo, N., & Tanabe, N. (2008). Low back pain in childhood and adolescence: A cross sectional study in Niigata City. *Eur Spine J*, *17*(11), pp.1441-1447.
39. Sjölie, A. N. (2004). Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scand J Med Sci Sports*, *14*(3), pp.168-175.
40. Standaert, C. J., & Herring, S. A. (2000). Spondylolysis: A critical review. *Br J Sports Med*, *34*(6), 415-422.
41. Wells, K. F., & Dillon, E. K. (1952). The sit-and-reach. A test of back and leg flexibility. *The Res Quarterly*, *23*, pp.115-118.

4. EFFECTIVENESS OF PHYSICAL EDUCATION-BASED INTERVENTION PROGRAMS FOR DEVELOPING AND MAINTAINING HEALTH-RELATED PHYSICAL FITNESS IN SCHOOLCHILDREN (PAPERS IX-XVI)

Paper IX

LAS CLASES DE EDUCACIÓN FÍSICA SOLO MEJORAN LA CAPACIDAD CARDIORESPIRATORIA DE LOS ALUMNOS CON MENOR CONDICIÓN FÍSICA: UN ESTUDIO DE INTERVENCIÓN CONTROLADO

[PHYSICAL EDUCATION CLASSES ONLY IMPROVE CARDIORESPIRATORY FITNESS OF STUDENTS WITH LOWER PHYSICAL FITNESS: A CONTROLLED INTERVENTION STUDY]

Mayorga-Vega, D., & Viciana, J.

Nutrición Hospitalaria

2015, *32*(1), 330-335.

Nutrición Hospitalaria

Nutr Hosp. 2015;32(1):330-335
ISSN 0212-1611 • CODEN NUHOEQ
S.V.R. 318

Original/*Deporte y ejercicio*

Las clases de educación física solo mejoran la capacidad cardiorrespiratoria de los alumnos con menor condición física: un estudio de intervención controlado

Daniel Mayorga-Vega y Jesús Viciana

Departamento de Educación Física y Deportiva, Universidad de Granada, España.

Resumen

Introducción: **los profesores de educación física deben llevar a cabo programas de intervención con el fin de mejorar los niveles de capacidad cardiorrespiratoria de sus alumnos. Sin embargo, debido a la baja carga lectiva, dichos programas podrían no mejorar la capacidad cardiorrespiratoria de los estudiantes con mayores niveles iniciales.**

Objetivos: **el objetivo del presente estudio fue comparar el efecto de un programa de acondicionamiento físico realizado durante las dos clases de educación física sobre la capacidad cardiorrespiratoria en función del nivel de condición física inicial de los estudiantes.**

Métodos: **una muestra de 71 estudiantes de educación primaria y 107 de educación secundaria obligatoria fue asignada aleatoriamente (por clases naturales) al grupo control y experimental. Durante las clases de educación física, los estudiantes experimentales realizaron un programa de acondicionamiento físico.**

Resultados: **los resultados del análisis de varianza de un factor mostraron que los estudiantes experimentales con menor nivel de condición física incrementaron estadísticamente su capacidad cardiorrespiratoria con respecto a los estudiantes controles ($p < 0,01$). En cambio, no se encontraron diferencias estadísticamente significativas entre los estudiantes experimentales con mayor nivel y los controles ($p > 0,05$).**

Conclusiones: **durante las clases de educación física, solo los estudiantes con menor condición física parecen incrementar su capacidad cardiorrespiratoria. En cambio, los alumnos con mayor nivel de condición física no se benefician de estos programas. Por ello, con el objetivo real de incrementar la capacidad cardiorrespiratoria de todos los jóvenes, parece necesario aumentar la carga lectiva de la asignatura de educación física.**

(*Nutr Hosp*. 2015;32:330-335)

DOI:10.3305/nh.2015.32.1.8919

Palabras clave: *Escuela. Condición física. Resistencia cardiovascular. Niños. Adolescentes.*

PHYSICAL EDUCATION CLASSES ONLY IMPROVE CARDIORESPIRATORY FITNESS OF STUDENTS WITH LOWER PHYSICAL FITNESS: A CONTROLLED INTERVENTION STUDY

Abstract

Introduction: **physical education teachers are required to carry out intervention programs in order to improve students´ cardiorespiratory fitness levels. Nevertheless, due to the low academic load, such programs may not improve cardiorespiratory fitness of students with higher baseline levels.**

Objectives: **the purpose of the present study was to compare the effect of a physical fitness program conducted during the two physical education classes on the cardiorespiratory fitness depending on the students´ physical fitness baseline levels.**

Methods: **a sample of 71 primary school students and 107 secondary school students was randomly assigned (by natural groups) to the control and experimental groups. During physical education classes, experimental students performed a physical fitness program.**

Results: **the analysis of variance results showed that the experimental students with lower physical fitness levels improved statistically significantly their cardiorespiratory fitness comparing with control students ($p < 0.01$). However, no statistically significant differences between the experimental students with higher levels and control students were found ($p > 0.05$).**

Conclusions: **during physical education classes, only students with lower physical fitness levels seem to improve cardiorespiratory fitness. On the other hand, students with higher physical fitness levels do not benefit from these physical education-based programs. Therefore, in order to improve cardiorespiratory fitness of all young people, it seems necessary to increase the academic load of the physical education subject.**

(*Nutr Hosp*. 2015;32:330-335)

DOI:10.3305/nh.2015.32.1.8919

Key words: *School. Physical fitness. Physical endurance. Children. Adolescents.*

Correspondencia: Daniel Mayorga-Vega.
Facultad de Ciencias del Deporte. Departamento de Educación Física y Deportiva. Universidad de Granada.
Carretera de Alfacar, s/n, 18011, Granada (España).
E-mail: dmayorgavega@gmail.com

Recibido: 5-III-2015.
Aceptado: 8-IV-2015.

Abreviaturas

EF: Educación física.
GC: Grupo control.
GE1: Grupo experimental 1 (estudiantes experimentales con menor capacidad cardiorrespiratoria).
GE2: Grupo experimental 2 (estudiantes experimentales con mayor capacidad cardiorrespiratoria).
ANOVA: Análisis de la varianza.

Introducción

En la actualidad la capacidad cardiorrespiratoria es considerada uno de los principales marcadores de salud entre los adultos[1], incluso por encima de otros marcadores tradicionales como el peso corporal, la presión arterial o el nivel de colesterol[2]. Además, hoy en día hay una fuerte evidencia científica que indica que mayores niveles de capacidad cardiorrespiratoria durante la infancia y la adolescencia se asocian con un menor riesgo de desarrollar síndrome metabólico, así como con un perfil cardiovascular más saludable durante la edad adulta[3]. Entre los jóvenes mejorar la capacidad cardiorrespiratoria además parece tener efectos positivos sobre marcadores psicológicos de salud tales como depresión, ansiedad, estado de ánimo y autoestima, así como también parece estar asociado con un mayor rendimiento académico[4].

Desafortunadamente, en las últimas décadas existe un importante descenso de los niveles de capacidad cardiorrespiratoria entre los jóvenes[5]. En la actualidad las cifras son especialmente alarmantes entre los jóvenes españoles, ya que uno de cada cinco adolescentes posee unos niveles de capacidad cardiorrespiratoria indicativo de riesgo cardiovascular futuro[6]. Por tanto, en España las políticas de promoción de salud deberían estar diseñadas también para promover la mejora de los niveles de capacidad cardiorrespiratoria desde la infancia y adolescencia[4]. En muchas circunstancias, las escuelas pueden jugar un papel importante en este problema de salud pública[4]. En esta línea, por ejemplo, los profesores españoles de educación física (EF) deben llevar a cabo programas de intervención con el fin de mejorar los niveles de capacidad cardiorrespiratoria de sus estudiantes[7,8].

Sin embargo, hoy en día los profesores de EF deben hacer frente a diferentes limitaciones para lograr un adecuado desarrollo de los niveles de capacidad cardiorrespiratoria de sus alumnos[9,10]. Por un lado, el profesor de EF debe desarrollar numerosos contenidos curriculares en cada curso escolar[7,8]. Por ejemplo, en educación secundaria existen cuatro grandes bloques de contenidos que los profesores de EF tienen que desarrollar: Condición física y salud, juegos y deportes, expresión corporal y actividades en el medio natural[8]. Por otro lado, la asignatura de EF se encuentra restringida por su limitada asignación de tiempo curricular[7,8]. Como en la mayoría de países de nuestro entorno[11], en el sistema educativo español sólo hay dos clases a la semana dedicada a la EF[7,8]. Además, su distribución depende de criterios de horarios no relacionados con la actividad física. Por tanto, en ocasiones las dos clases de EF se podrían impartir en dos días consecutivos sin dejar un mayor tiempo de descanso, en la primera hora de la mañana cuando los alumnos acaban de tomar el desayuno o al final de la mañana cuando gran parte del año las temperaturas son elevadas[10].

A pesar de sus numerosas limitaciones, estudios previos han encontrado que un programa de acondicionamiento físico durante las clases de EF dos veces por semana puede mejorar la capacidad cardiorrespiratoria de los escolares[12]. Además, otros estudios han analizado el efecto positivo de incrementar el volumen mediante un mayor número de sesiones semanales[12-17], la intensidad de las sesiones[18], o ambos de manera conjunta[13]. Sin embargo, a pesar de que el incremento de la capacidad cardiorrespiratoria está negativamente asociada a la condición física inicial[19], los estudios que analiza la influencia del nivel inicial de capacidad cardiorrespiratoria de los estudiantes sobre el efecto de la intervención son escasos[15], no habiéndose encontrado estudios previos en EF con solo dos sesiones semanales. Debido a la baja frecuencia semanal de la asignatura, quizás solo los alumnos con menor nivel de capacidad cardiorrespiratoria serán los únicos beneficiados de estos programas en EF. Consecuentemente, el objetivo del presente estudio fue comparar el efecto de un programa de acondicionamiento físico realizado durante las dos clases de EF sobre la capacidad cardiorrespiratoria en función del nivel de condición física inicial de los estudiantes.

Métodos

Participantes

El presente estudio está realizado mediante dos muestras independientes. La primera muestra estuvo formada por 71 estudiantes (32 niños y 39 niñas) de cuatro clases de 6º curso de educación primaria. La segunda muestra estuvo constituida por 107 estudiantes (68 niños y 39 niñas) de seis clases de 2º curso de educación secundaria obligatoria. Por razones prácticas y debido a la naturaleza del presente estudio, se empleó un diseño *cluster-randomized controlled trial*[20-22]. Para cada muestra, las clases naturales se asignaron aleatoriamente al grupo control (GC) y el grupo experimental. Posteriormente, cada grupo experimental fue dividido a su vez en dos subgrupos según los niveles de capacidad cardiorrespiratoria basales de sus estudiantes (GE1 $< P_{50}$ y GE2 $\geq P_{50}$) (Tabla I).

Todos los participantes estaban libres de padecer cualquier trastorno de salud que le impidieran realizar actividad física como, por ejemplo, enfermedades del corazón, asma no controlada o problemas osteoarticulares. El criterio de inclusión fue tener una asistencia

Tabla I
Características generales de los participantes estudiados

	Control (M ± DE)	*Experimental 1 (M ± DE)*	*Experimental 2 (M ± DE)*
Educación primaria	*(n = 36)*	*(n = 17)*	*(n = 18)*
Edad (años)	11,1 ± 0,4	11,2 ± 0,4	10,9 ± 0,4
Masa corporal (kg)	44,9 ± 11,2	43,3 ± 10,5	39,4 ± 7,9
Talla (cm)	148,9 ± 5,8	144,7 ± 8,6	145,3 ± 5,3
Índice de masa corporal (kg/m^2)	20,2 ± 4,4	20,4 ± 3,3	18,6 ± 3,2
Educación secundaria	*(n = 45)*	*(n = 31)*	*(n = 31)*
Edad (años)	12,5 ± 0,7	12,4 ± 0,6	12,5 ± 0,6
Masa corporal (kg)	52,7 ± 14,1	54,2 ± 11,5	49,9 ± 9,5
Talla (cm)	158,9 ± 7,4	156,2 ± 7,4	159,4 ± 8,0
Índice de masa corporal (kg/m^2)	20,7 ± 4,4	22,1 ± 3,7	19,5 ± 2,7

Nota. M = media; DE = desviación estándar.

a las clases de EF del programa de intervención igual o superior al 85%. Todos los estudiantes y sus tutores legales fueron plenamente informados acerca de las características del estudio y firmaron un consentimiento informado. El protocolo del estudio fue aprobado por el Comité Ética de la Universidad de Granada, así como que respetaba el acuerdo vigente de la Declaración de Helsinki sobre normas éticas en investigación.

Medidas

La evaluación se llevó a cabo durante las clases de EF al comienzo y al final del programa de intervención (pre-intervención y post-intervención, respectivamente). Cada evaluación fue realizada por el mismo evaluador, instrumentos y condiciones. Las medidas fueron tomadas en una instalación deportiva cubierta con suelo antideslizante, en las mismas condiciones medioambientales, y a la misma hora y día de la semana para cada estudiante. Además, antes de la evaluación todos los participantes realizaron un calentamiento estandarizado de cinco minutos de carrera de baja a moderada intensidad.

Para evaluar la capacidad cardiorrespiratoria de los alumnos se empleó el test Course Navette[23]. Brevemente, los participantes corrieron entre dos líneas paralelas a 20 m de distancia al ritmo marcado por un sonido grabado. El test comenzaba a una velocidad de 8,5 km/ h e incrementaba 0,5 km/ h cada minuto. Con los estudiantes de educación primaria, un evaluador corrió junto con los escolares para ayudarles a mantener la velocidad adecuada. La prueba terminaba cuando los participantes dejaban de correr debido a la fatiga o cuando no lograban llegar a la línea antes de la siguiente señal en dos ocasiones consecutivas. Cada participante realizó la prueba una vez y se retuvo el tiempo total en segundos. El test Course Navette ha demostrado unos valores adecuados de fiabilidad y validez entre los niños y adolescentes[23,24].

Procedimiento

Una descripción detallada del programa intervención ha sido publicada previamente[25,26]. Brevemente, a los estudiantes experimentales se aplicó un programa de acondicionamiento físico dos veces por semana durante sus clases de EF. El programa de intervención tuvo una duración de ocho semanas para los estudiantes de educación primaria y de nueve semanas para los estudiantes de educación secundaria. Sin embargo, debido a que dos sesiones de cada programa coincidieron con días festivos, los estudiantes experimentales de educación primaria y secundaria completaron un total de 14 y 16 sesiones, respectivamente. Cada sesión de intervención duró 50 minutos aproximadamente y consistió en un período de 5 a 10 minutos de calentamiento, 35 a 40 minutos de parte principal, y cinco minutos de vuelta a la calma. Se puso especial énfasis en alcanzar una intensidad moderada a vigorosa durante las sesiones experimentales.

Por su parte, los estudiantes del GC también participaron en dos sesiones de EF por semana. Además, las sesiones también tuvieron una duración y estructura temporal similar a la de los estudiantes experimentales. Sin embargo, los contenidos y metodologías seguidas durante las sesiones fueron diferentes. En cuanto a los contenidos, los estudiantes del GC realizaron sesiones del bloque juegos y deportes en lugar de condición física. Concretamente, los estudiantes del GC de educación primaria realizaron sesiones de juegos tradicionales y los estudiantes de educación secundaria voleibol y bádminton. Por otro lado, en cuanto a la metodología, se puso un especial énfasis al aprendizaje técnico-táctico en vez de a la intensidad de la tarea.

Análisis estadístico

Se realizó una estadística descriptiva (medias ± desviaciones estándar) de la edad, masa corporal, talla, índice de masa corporal, y los valores de la capacidad cardiorrespiratoria. Se utilizó el análisis de la varianza (ANOVA) de un factor para estudiar las posibles diferencias en las características generales entre los grupos de cada muestra. Posteriormente, el efecto del programa de intervención se estudió mediante un ANOVA de un factor, incluyendo grupo como factor fijo y cambio pre-post intervención como variable dependiente. Luego, se realizaron comparaciones por pares (*post hoc*) con la corrección de Bonferroni. A continuación, el tamaño del efecto *g* de Hedges se utilizó para estimar la magnitud del efecto del programa de intervención[27]. Por último, la fiabilidad test-retest de la medida se examinó mediante el coeficiente de correlación intraclase del ANOVA de dos factores con el intervalo de confianza al 95%[28]. Todos los análisis estadísticos se realizaron mediante el paquete estadístico SPSS versión 20.0 para Windows (IBM® SPSS® Statistics 20). El nivel de significación estadística se estableció en $p < 0{,}05$.

Resultados

Los estudiantes experimentales obtuvieron una asistencia media del 94% y 97% para la muestra de educación primaria y secundaria, respectivamente. En cuanto a las características generales de los participantes, para cada muestra los resultados del ANOVA de un factor no mostraron diferencias estadísticamente significativas entre los grupos estudiados ($p > 0{,}05$), excepto en el índice de masa corporal para la muestra de educación secundaria ($p < 0{,}05$). Por otro lado, el coeficiente de correlación intraclase para la medida de capacidad cardiorrespiratoria mostró unos valores adecuados: 0,90 (0,81-0,95) y 0,96 (0,93-0,98) para la muestra de educación primaria y secundaria, respectivamente.

La tabla II muestra el efecto del programa de intervención sobre la capacidad cardiorrespiratoria. Los resultados del ANOVA de un factor sobre los valores obtenidos en la prueba de Course Navette mostraron un efecto estadísticamente significativo para tanto para la muestra de educación primaria [$F(2, 68) = 15{,}335$; $p < 0{,}001$; $\eta^2_p = 0{,}228$; P = 0,999] como de educación secundaria [$F(2, 104) = 7{,}633$; $p = 0{,}001$; $\eta^2_p = 0{,}183$; P = 0,938]. Posteriormente, las comparaciones por pares con la corrección de Bonferroni mostraron que los estudiantes del GE1 incrementaron estadísticamente su capacidad cardiorrespiratoria con respecto a los estudiantes del GC ($p < 0{,}01$). En cambio, no se encontraron diferencias estadísticamente significativas entre los estudiantes del GE2 y GC ($p > 0{,}05$). Por otro lado, los estudiantes de secundaria del GE1 mostraron un incremento estadísticamente significativo con respecto al GE2 ($p < 0{,}001$), así como que entre los estudiantes de primaria hubo una tendencia a la significación ($p < 0{,}10$).

Discusión

El objetivo del presente estudio fue comparar el efecto de un programa de acondicionamiento físico realizado durante las dos clases de EF sobre la capacidad cardiorrespiratoria en función del nivel de condición física inicial de los estudiantes. Entre otras muchas tareas, los profesores de EF deben llevar a cabo

Tabla II
Efecto de la intervención sobre la capacidad cardiorrespiratoria (Course Navette, s)

Grupo	*Pre-intervención (M ± DE)*	*Post-intervención (M ± DE)*	*Diferencia (M ± DE)*	p^a	*Tamaño del efecto*[b]	
Educación primaria						
GE1 ($n = 17$)	78,1 ± 24,2	139,9 ± 47,0	61,8 ± 55,2**	0,001	GE1-GC	0,74
GE2 ($n = 18$)	239,5 ± 62,0	255,2 ± 102,4	15,7 ± 74,4		GE2-GC	0,26
GC ($n = 36$)	193,9 ± 97,5	184,5 ± 96,2	- 9,4 ± 58,0		GE1-GE2	0,48
Educación secundaria						
GE1 ($n = 31$)	216,4 ± 54,3	252,9 ± 76,0	36,5 ± 35,3***	< 0,001	GE1-GC	0,44
GE2 ($n = 31$)	416,5 ± 84,2	407,5 ± 91,4	- 9,0 ± 46,9		GE2-GC	0,07
GC ($n = 45$)	294,3 ± 121,7	276,5 ± 116,0	- 17,8 ± 46,1†††		GE1-GE2	0,37

Nota. M = media; DE = desviación estándar; GE1 = grupo experimental 1; GE2 = grupo experimental 2; GC = grupo control.
[a] Nivel de significación del análisis de varianza de un factor con las comparaciones por pares (*post hoc*) con la corrección de Bonferroni: Cambio estadísticamente significativo del GC-GE1 (** $p < 0{,}01$, *** $p < 0{,}001$) y del GE1-GE2 (††† $p < 0{,}001$).
[b] Tamaño del efecto *g* de Hedges.

programas de intervención con el fin de mejorar los niveles de capacidad cardiorrespiratoria de sus estudiantes[7,8]. Para ello, hoy en día los profesores de EF deben hacer frente a diferentes limitaciones[9,10]. Por ejemplo, al igual que en la mayoría de países de nuestro entorno[11], en España la asignatura de EF se encuentra limitada a solo dos clases a la semana[7,8].

Algunos estudios previos han encontrado que un programa de acondicionamiento físico durante las clases de EF dos veces por semana mejora la capacidad cardiorrespiratoria de los escolares[12]. Sin embargo, se ha encontrado que el efecto de estos programas sobre el incremento de la capacidad cardiorrespiratoria está negativamente asociado a la condición física inicial de los niños[19]. Por tanto, se planteó la hipótesis de que, debido a la baja frecuencia semanal de la asignatura de EF, quizás solo los alumnos con bajo nivel serían los únicos beneficiados de estos programas. En este sentido, los resultados del presente estudio mostraron cómo solo los estudiantes con menor condición física, tanto de educación primaria y como de secundaria, incrementaron estadísticamente su capacidad cardiorrespiratoria. En cambio, la intervención no incrementó la capacidad cardiorrespiratoria de aquellos los alumnos con mayor condición física inicial.

De manera similar al presente estudio, Resaland et al.[15] examinaron el efecto de un programa diario de actividad física sobre la capacidad cardiorrespiratoria en escolares de educación primaria según su condición física basal. Estos autores encontraron cómo solo los estudiantes con baja y moderada condición física basal (cuartiles 1, 2 y 3) presentaron mejoras estadísticamente significativas tras la aplicación del programa. En cambio, la intervención no incrementó la capacidad cardiorrespiratoria de aquellos alumnos con mayor condición física inicial (cuarto cuartil). Debido a la mayor frecuencia del estudio anterior con respecto a la asignatura de EF de nuestro país (es decir, cinco sesiones a la semana), parecía necesario el presente estudio con el objeto de comprobar estos hallazgos en el contexto de la EF escolar española (es decir, solo dos sesiones a la semana).

En cuanto a la magnitud del efecto de la intervención, después del programa de acondicionamiento físico el tamaño del efecto del presente estudio fue moderado/alto para los estudiantes con menor condición física y trivial/bajo para los estudiantes con mayor condición física. Estos hallazgos indican que el programa de intervención del presente estudio fue solo efectivo para aquellos alumnos con menor nivel. De modo similar, aunque Resaland et al.[15] encontraron un efecto moderado/alto para los estudiantes con menor condición física (g = 0,49, 0,70 y 0,83 para los cuartiles 1, 2 y 3, respectivamente), el efecto fue bajo para los estudiantes con alta condición física (g = 0,23). Por otro lado, aunque en el presente estudio el tamaño del efecto fue considerablemente mayor para los estudiantes de educación primaria que para los de educación secundaria, debido al hecho de que los programas de intervención fueron diferentes para cada muestra, no podríamos afirmar si las diferencias de edad entre ambas muestras podrían contribuir a dichas diferencias.

En conclusión, de lo que conocemos el presente estudio es el primero en comparar el efecto de un programa de acondicionamiento físico durante las dos clases de EF sobre la capacidad cardiorrespiratoria en función del nivel de condición física inicial de los estudiantes. Los resultados de este estudio sugieren que durante las clases de EF, tanto de educación primaria como de educación secundaria, solo los estudiantes con menor condición física pueden incrementar su capacidad cardiorrespiratoria. En cambio, los alumnos con mayor nivel de condición física no se beneficiarían de estos programas escolares. Por ello, con el objetivo real de incrementar la capacidad cardiorrespiratoria de todos los jóvenes, parece necesario aumentar la carga lectiva de la asignatura de EF, por ejemplo, mediante un aumento de la frecuencia semanal.

Agradecimientos

Agradecemos a todos los alumnos y padres por su participación voluntaria e incondicional en el presente estudio. Agradecemos también a todos los miembros del centro escolar por su entusiasmo y colaboración, especialmente a los directores y profesores de educación física. El primer autor recibe una ayuda del programa de Formación del Profesorado Universitario (FPU) del Ministerio de Educación, Cultura y Deporte (AP2010-5905).

Conflicto de intereses

Los autores declaramos no tener ningún conflicto de interés.

Referencias

1. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* 2009;301(19):2024-35.
2. Blair SN. Physical inactivity: The biggest public health problem of the 21st century. *Br J Sports Med* 2009;43(1):1-2.
3. Ruiz JR, Castro-Pinero J, Artero EG, Ortega FB, Sjöström M, Suni J, et al. Predictive validity of health-related fitness in youth: A systematic review. *Br J Sports Med* 2009;43(12):909-23.
4. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: A powerful marker of health. *Int J Obes (Lond)* 2008;32(1):1-11.
5. Tomkinson GR, Olds TS. Secular changes in pediatric aerobic fitness test performance: The global picture. *Med Sport Sci* 2007;50:46-66.
6. Ortega FB, Ruiz JR, Castillo MJ, Moreno LA, González-Gross M, Wärnberg J, et al. Bajo nivel de forma física en los adolescentes españoles. Importancia para la salud cardiovascular futuro (Estudio AVENA). *Rev Esp Cardiol* 2005;58(8):898-909.
7. Ministerio de Educación y Ciencia. Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mí-

nimas correspondientes de la educación primaria. *Boletín Oficial del Estado* 2006;293:43053-102.
8. Ministerio de Educación y Ciencia. Real Decreto 1613/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la educación secundaria obligatoria. *Boletín Oficial del Estado* 2007;5:677-773.
9. Viciana J, Mayorga-Vega D, Cocca A. Modelo de aprendizaje exitoso en educación física y su mantenimiento. Estudio del efecto del refuerzo intermitente sobre la condición física. *Rev Iberoam Psicol Ejer Deporte* 2014;9(1):155-71.
10. Viciana J, Mayorga-Vega D, Merino-Marban R. Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In: Todaro R (ed) Handbook of physical education research. Role of school programs, children´s attitudes and health implications. New York: *Nova Science Publisher* 2014:237-52.
11. European Commission/ EACEA/ Eurydice. Physical education and sport at school in Europe Eurydice Report. Luxembourg: *Publications Office of the European Union* 2013.
12. Ramírez Lechuga J, Muros Molina JJ, Morente Sánchez J, Sánchez Muñoz C, Femia Marzo P, Zabala Díaz M. Efecto de un programa de entrenamiento aerobic de 8 semanas durante las clases de educación física en adolescentes. *Nutr Hosp* 2012;27(3):747-54.
13. Ardoy DN, Fernández-Rodríguez JM, Ruiz JR, Chillón P, España-Romero V, Castillo MJ, et al. Mejora de la condición física en adolescentes a través de un programa de intervención educativa: Estudio EDUFIT. *Rev Esp Cardiol* 2011; 64(6): 481-91.
14. Kriemler S, Zahner L, Schindler C, Meyer U, Hartmann T, Hebestreit H, et al. Effect of school based physical activity programme (KISS) on fitness and adiposity in primary schoolchildren: Cluster randomised controlled trial. *BMJ* 2010;340:785.
15. Resaland GK, Andersen LB, Mamen A, Anderssen SA. Effects of a 2-year school- based daily physical activity intervention on cardiorespiratory fitness: The Sogndal school-intervention study. *Scand J Med Sci Sports* 2011;21(2):302-9.
16. Sollerhed AC, Ejlertsson G. Physical benefits of expanded physical education in primary school: findings from a 3-year intervention study in Sweden. *Scand J Med Sci Sports* 2008;18(1):102-7.
17. Walther C, Gaede L, Adams V, Gelbrich G, Leichtle A, Erbs S, et al. Effect of increased exercise in school children on physical fitness and endothelial progenitor cells: A prospective randomized trial. *Circulation* 2009;120:2251-9.
18. Baquet G, Berthoin S, Gerbeaux M, Van Praagh E. High-intensity aerobic training during a 10 week one-hour physical education cycle: Effects on physical fitness of adolescents aged 11 to 16. *Int J Sports Med* 2001;120(22):295-300.
19. Eliakim A, Barstow TJ, Brasel JA, Ajie H, Lee WN, Renslo R, et al. Effect of exercise training on energy expenditure, muscle volume, and maximal oxygen uptake in female adolescents. *J Pediatr* 1996;129(4):537-43.
20. Mayorga-Vega D, Merino-Marban R, Garrido FJ, Viciana J. Comparison between warm-up and cool-down stretching programs on hamstring extensibility gains in primary schoolchildren. *Phys Act Rev* 2014;2:16-24.
21. Mayorga-Vega D, Merino-Marban R, Vera-Estrada F, Viciana J. Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology* 2014;46(2):227-33.
22. Merino-Marban R, Mayorga-Vega D, Fernandez-Rodriguez E, Vera Estrada F, Viciana J. Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *Eur Physic Educ Rev* 2015;21(1):83-92.
23. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 meter shuttle run test for aerobic fitness. *J Sports Sci* 1988;6(2):93-101.
24. Mayorga-Vega D, Aguilar-Soto P, Viciana J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *J Sports Sci Med* 2015;14(3):536-47.
25. Mayorga-Vega D, Viciana J, Cocca A. Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *J Hum Kinet* 2013;37:153-60.
26. Mayorga-Vega D, Montoro-Escaño J, Merino-Marban R, Viciana J. Effects of a physical education-based program on health-related physical fitness and its maintenance in high school students: A cluster-randomized controlled trial. *Eur Physic Educ Rev* In press.
27. Hedges LV. Effect sizes in cluster-randomized designs. *J Educ Behav Stat* 2007;32(4):341-70.
28. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420-28.

Paper X

A PHYSICAL EDUCATION-BASED STRETCHING PROGRAM PERFORMED ONCE A WEEK ALSO IMPROVES HAMSTRING EXTENSIBILITY IN SCHOOLCHILDREN: A CLUSTER-RANDOMIZED CONTROLLED TRIAL

Mayorga-Vega, D., Merino-Marban, R., Real, J. & Viciana, J.

Nutrición Hospitalaria

2015, *32*(4), 1715-1721.

Nutrición Hospitalaria

Nutr Hosp. 2015;32(4):1715-1721
ISSN 0212-1611 • CODEN NUHOEQ
S.V.R. 318

Original/*Deporte y ejercicio*

A physical education-based stretching program performed once a week also improves hamstring extensibility in schoolchildren: a cluster-randomized controlled trial

Daniel Mayorga Vega[1], Rafael Merino Marbán[2], Javier Real[3] and Jesús Viciana[1]

[1]*Department of Physical Education and Sport, University of Granada.* [2]*Facultad de Ciencias de la Educación, Universidad de Málaga (UMA).* [3]*Servicio de Deportes, Universidad Pablo de Olavide. Spain.*

Abstract

Introduction: **physical education teachers are required to carry out intervention programs for students to achieve health-enhancing flexibility levels. Unfortunately, to our knowledge, there are no studies examining the effect of a stretching program carried out only once a week on schoolchildren.**

Objectives: **the purpose of the present study was to compare the effects of a short-term stretching intervention program performed once and twice a week on hamstring extensibility among schoolchildren in the physical education setting.**

Methods: **a sample of 180 high school students aged 12-14 years old was randomly assigned (by natural groups) to a control group, experimental group 1 and experimental group 2. During physical education classes, experimental group students performed a stretching program for eight weeks. The experimental group 1 and 2 performed the stretching program once and twice a week, respectively.**

Results: **the analysis of variance results showed that the students of both experimental groups improved statistically significantly their hamstring extensibility when compared with the control group students ($p < 0.01$). Nevertheless, no statistically significant differences between the two experimental groups were found ($p > 0.05$).**

Conclusions: **a short-term stretching program performed only once a week improves hamstring extensibility in schoolchildren. When the stretching program is performed twice a week, the improvement in students´ hamstring extensibility is not statistically higher.**

(*Nutr Hosp*. 2015;32:1715-1721)

DOI:10.3305/nh.2015.32.4.9302

Key words: *School. Range of motion. Static stretching. Physical fitness. Adolescents.*

Correspondence: Daniel Mayorga Vega.
Facultad de Ciencias del Deporte, Departamento de Educación Física y Deportiva, Universidad de Granada, Carretera de Alfacar, s/n, 18011 (Granada), España.
E-mail: dmayorgavega@gmail.com

Recibido: 22-V-2015.
Aceptado: 26-VI-2015.

UN PROGRAMA DE ESTIRAMIENTO EN EDUCACIÓN FÍSICA REALIZADO UNA VEZ A LA SEMANA TAMBIÉN MEJORA LA EXTENSIBILIDAD ISQUIOSURAL EN ESCOLARES: UN DISEÑO CONTROLADO ALEATORIO POR GRUPOS

Resumen

Introducción: **los profesores de educación física deben llevar a cabo programas de intervención para que los estudiantes alcancen niveles de flexibilidad saludables. Lamentablemente, no se conocen estudios que examinen el efecto de un programa de estiramiento llevado a cabo solo una vez por semana en escolares.**

Objetivos: **el objetivo del presente estudio fue comparar los efectos de un programa de intervención de estiramiento a corto plazo realizado una y dos veces por semana sobre la extensibilidad isquiosural en escolares en el contexto de la educación física.**

Métodos: **una muestra de 180 estudiantes de educación secundaria de 12-14 años fue asignado aleatoriamente (por grupos naturales) a los grupos control, experimental 1 y experimental 2. Durante las clases de educación física, los estudiantes experimentales realizaron un programa de estiramiento durante ocho semanas. El grupo experimental 1 y 2 realizaron el programa de estiramiento una y dos veces por semana, respectivamente.**

Resultados: **los resultados del análisis de varianza mostraron que los estudiantes de ambos grupos experimentales obtuvieron una mejora estadísticamente significativa de la extensibilidad isquiosural, comparado con los estudiantes del grupo de control ($p < 0,01$). Sin embargo, no se encontraron diferencias estadísticamente significativas entre los dos grupos experimentales ($p > 0,05$).**

Conclusiones: **un programa de estiramiento a corto plazo realizado una vez por semana mejora la extensibilidad isquiosural en escolares. Cuando el programa de estiramiento se lleva a cabo dos veces por semana, la mejora en la extensibilidad isquiosural de los estudiantes no es estadísticamente mayor.**

(*Nutr Hosp*. 2015;32:1715-1721)

DOI:10.3305/nh.2015.32.4.9302

Palabras clave: *Escuela. Rango de movimiento. Estiramiento estático. Condición física. Adolescentes.*

Abbreviations

PE: Physical education.
CG: Control group.
EG1: Experimental group 1 (students that performed the intervention once a week).
EG2: Experimental group 2 (students that performed the intervention twice a week).
SR: Sit-and-reach.
ANOVA: Analysis of variance.

Introduction

Hamstring extensibility is a physical fitness component widely recognized as an important marker of health and quality of life[1]. Hamstring extensibility plays an important role in protecting the spine from potential disorders[2] and thereby allowing people to execute their normal daily activities[3]. Particularly, young people with adequate hamstring extensibility seem to have a lower risk of current low back pain[4,5] and neck tension[6], as well as a lower risk of low back pain later during adulthood[7,8].

Unfortunately, hamstring extensibility decreases significantly during youth[9]. Moreover, nowadays the number of young people with shortened hamstring extensibility is particularly alarming, since over one in five Spanish schoolchildren has a hamstring extensibility level indicative of health risk[10]. Therefore, in Spain health promotion policies should also be designed to promote the achievement of health-enhancing levels among childhood and adolescence[11]. In many circumstances, schools could play an important role in combating this important public health issue[11]. In this line, for instance, Spanish physical education (PE) teachers are required to carry out intervention programs for students to achieve health-enhancing flexibility levels[12].

Previous studies have found that a PE-based stretching program carried out twice a week improves hamstring extensibility in schoolchildren[13-15]. Additionally, other studies have analyzed the positive effect of increasing the number of weekly sessions from two to four[16]. However, apart from the many curricular contents that PE teachers must teach each academic year, currently the subject is too restricted by its limited curriculum time allocation[17]. Therefore, since stretching programs cannot be allocated a large part of the PE time, including a one-session-per-week stretching program could be more feasible.

Unfortunately, to our knowledge there are no studies examining the effect of a stretching program carried out only once a week in young people, the number of related studies on adults is scarce as well[18]. Consequently, the purpose of the present study was to compare the effects of a short-term stretching intervention program performed once and twice a week on hamstring extensibility among schoolchildren in a PE setting.

Methods

Participants

A sample of 180 students, 94 boys and 86 girls, aged 12-14 years old from six different first/ second-grade PE classes at a private-public high school center participated in the present study. For practical reasons and due to the nature of the present study (i.e., intervention focused on natural groups in a school setting) a cluster randomized controlled design was used[14,19]. The six natural classes balanced by grade were assigned randomly to form one of the following study groups: a control group (CG), experimental group 1 (EG1) and experimental group 2 (EG2).

All the participants were free of orthopedic disorders such as episodes of hamstring and/ or lumbar injuries, fractures, surgery or pain in the spine or hamstring and/ or lumbar muscles over the past six months[20]. The inclusion criteria were: (a) correctly performing the two flexibility evaluations, (b) and having an attendance rate of 90% or higher for PE classes during the intervention period. Adolescents and their legal guardians were fully informed about all the features of the study and were required to sign an informed consent document. The study protocol respected the current agreement of the Declaration of Helsinki on ethical principles for research involving human subjects and it was approved by the Ethical Committee of the University of Granada.

Measures

Evaluation was carried out during the PE classes at the beginning and at the end of the stretching intervention program (pre-intervention and post-intervention, respectively). Each evaluation was carried out by the same tester, instrument, and under the same conditions. Pre-intervention and post-intervention measures were taken in an indoor sports facility on the same day of the week and at the same time for each student. Prior to the test, the students completed a standardized warm-up consisting of three minutes of jogging.

The classic sit-and-reach (SR) test was used to estimate students' hamstring extensibility. The students were assessed in sportswear and barefoot. At the beginning of the test, the students stood in front of the box, sat with their hips flexed, knees extended and both hands on the top of the ruler. From this position, the adolescents had to bend their trunk forward slowly and progressively (without swings) in order to reach the furthest possible distance and to remain still for at least two seconds (the score 23 cm corresponded to the tangent of the feet; accuracy 0.1 cm). The students were allowed to perform the test twice, one minute apart, and then the average score in cm was recorded[21]. The SR test has demonstrated high reliability and adequate criterion-related validity among young people[22].

Procedures

A stretching intervention program was applied to the EGs during their PE sessions. The EG students performed a stretching development program four minutes per session for eight weeks. While the EG2 students performed the stretching intervention program twice a week, the EG1 students performed it only once a week. Since one lesson of the EG2 could not be used, in the end the participants of the EG1 and EG2 completed a total of 8 and 15 sessions, respectively. During each intervention session, the EG students performed static hamstring stretches during the cool-down period[23]. In the PE setting, the stretching intervention has been called "intermittent teaching unit"[24].

Each intervention session included two 30-second sets of four stretching exercises. Three bidopal exercises and one unipodal exercise were performed in each session (Fig. 1). In all the stretching exercises, the students were placed with their hips flexed and knees fully extended. The toes were pointed to the ceiling with no hip rotation. From this position, the students flexed forward at the hip, trying to maintain the spine in neutral position as much as possible until a gentle stretch was felt in the hamstrings. The stretched position was held gently until the end point of the range of motion was reached (i.e., stretch to the point of feeling tightness, but no pain). Once this position was achieved, the adolescents held it for 30 seconds.

All the students were urged to maintain their normal levels of physical activity outside the supervised setting during the research period. During the intervention program period, all the students participated in their standard PE sessions. However, the CG students did not perform stretching exercises and were not aware of the purpose of the study. Both the standard PE sessions and the stretching intervention programs were carried out by the same PE teacher of the participating center for all the groups.

Statistical analysis

Descriptive statistics (means ± standard deviations/ frequency) for age, gender, body mass, body height, body mass index, extra-curricular sport practitioners, and SR scores were calculated. A one-way analysis of variance (ANOVA) was conducted to examine potential differences between the three groups in terms of body mass, body height, body mass index,

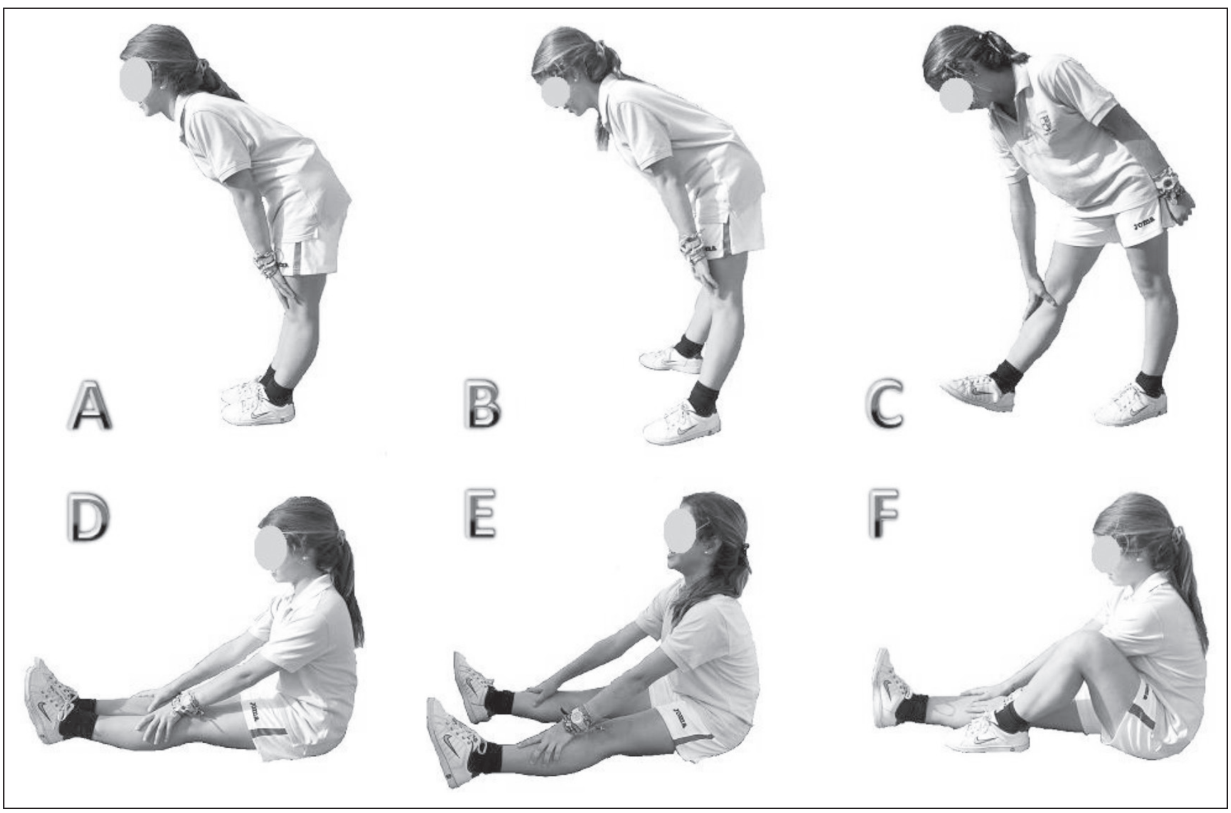


Fig. 1.—The six stretching exercises performed during the stretching intervention program: (a) standing with feet together; (b) standing with feet shoulder-width apart; (c) standing with only one leg extended; (d) sitting with feet together; (e) sitting with feet shoulder-width apart, and (f) sitting with only one leg extended. Session 1: (a), (b), (d) and (f); Session 2: (d), (e), (a) and (c).

and pre-intervention values of the SR test. Additionally, a chi-squared analysis was carried out to test the ratio differences of gender and extra-curricular sport practitioners between the three groups. Afterwards, the effect of the stretching intervention program on hamstring extensibility was examined using a one-way ANOVA, including *group* as a fixed factor (CG, EG1, EG2) and change pre-intervention - post-intervention as a dependent variable. Subsequently, the *post hoc* analyses with the Bonferroni adjustment were used for the pairwise comparisons. Moreover, the Hedges' g effect size was used to examine the magnitude of intervention effects[25]. The minimal detectable change was calculated in order to examine if the change score due to the intervention was true and reliable rather than due to measurement error[26]. Additionally, the test-retest reliability of the SR scores was estimated using the intraclass correlation coefficient from the two-way ANOVA[27], as well as the 95% confidence interval. All statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20). The statistical significance level was set at $p < 0.05$.

Results

Figure 2 shows the flow chart corresponding to the participants included in the present study. Although all of the 180 invited students agreed to participate, only 163 participants completed the two evaluations and attended 90% or more of the intervention program sessions. Table I shows the general characteristics of the participants studied and the pre/ post-intervention SR scores. The one-way ANOVA results did not show statistically significant differences in body mass, body height, body mass index, and SR pre-intervention values between groups ($p > 0.05$). Furthermore, the chi-square analyses showed that the three groups had a balanced representation of boys/ girls and extra-curricular sport practitioners/ non-practitioners ($p > 0.05$). The test-retest reliability of the SR scores was 0.997 (0.994-0.998).

Figure 3 shows the effect of the stretching intervention program on hamstring extensibility levels. The results of the one-way ANOVA on the average obtained in the SR test showed a statistically significant effect [$F(2, 160) = 11.192$; $p < 0.001$; $\eta^2_p = 0.123$; P = 0.991]. Subsequently, the pairwise comparisons with the Bonferroni adjustment showed that both EG1 and EG2 obtained a statistically significant improvement compared to the CG (EG1, $p = 0.005$; $g = 0.16$; EG2, $p < 0.001$; $g = 0.23$). Nevertheless, no statistically significant differences between the EG1 and EG2 were found ($p > 0.05$; $g = 0.07$). The minimal detectable change value of the SR score was 0.96 cm, when the average increase in the EG1 and EG2 was 1.41 and 1.93 cm, respectively.

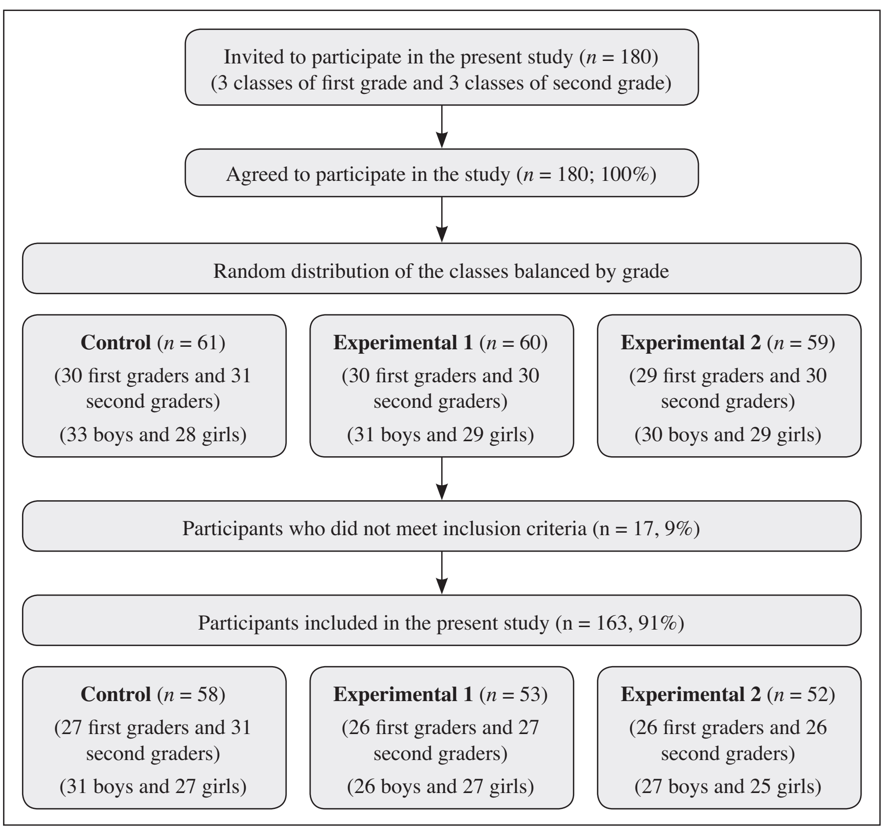


Fig. 2.—Flow chart corresponding to the participants included in the present study.

Table I
General characteristics (mean ± standard deviation/ frequency) of the participants and pre/ post-intervention sit-and-reach scores

	Control (n = *58)*	*Experimental 1* (n = *53)*	*Experimental 2* (n = *52)*
Age (years)	12.6 ± 0.6	12.7 ± 0.7	12.7 ± 0.6
Gender (boys/ girls)	31/ 27	26/ 27	27/ 25
Body mass (kg)	49.8 ± 8.1	47.3 ± 8.5	49.2 ± 9.7
Body height (cm)	158.3 ± 7.0	157.7 ± 7.5	156.4 ± 7.1
Body mass index (kg/ m^2)	19.8 ± 2.6	18.9 ± 2.2	20.0 ± 3.2
Extra-curricular sport (yes/ no)[a]	37/ 21	37/ 15	35/ 17
Pre-intervention SR score (cm)	20.4 ± 7.0	20.2 ± 6.7	20.7 ± 7.7
Post-intervention SR score (cm)	20.7 ± 7.4	21.7 ± 6.6	22.6 ± 8.2

Note. SR = Sit-and-reach.
[a] Children that regularly participated (yes) or not (no) at least twice per week in extra-curricular sport activities.

Discussion

The purpose of the present study was to compare the effects of a short-term stretching intervention program performed once and twice a week on hamstring extensibility among schoolchildren in the PE setting. Spanish PE teachers are required to carry out intervention programs for students to achieve health-enhancing flexibility levels[12]. Previous studies have found that a short-term stretching program carried out twice a week improves hamstring extensibility in schoolchildren[13-15]. However, apart from the many curricular contents that PE teachers must teach each academic year, currently the subject is too restricted by its limited curriculum time allocation[17]. Consequently, including a stretching program performs only once a week would be more feasible.

The results of the present study showed that a short-term stretching program performed only once a week improves hamstring extensibility in schoolchildren. Additionally, when the stretching program is performed twice a week, the improvement in students´ hamstring extensibility is not statistically different. To our knowledge there are no studies examining the effect of a one-session-per-week stretching program on schoolchildren. However, Santonja et al.[16] compared the effects of a stretching program performed two and four times a week on hamstring extensibility in schoolchildren. Contrary to the present results, these authors found that children performing a stretching program four sessions per week obtained statistically greater improvements on hamstring extensibility compared with two sessions per week. However, apart from the difference in frequency, the fact that

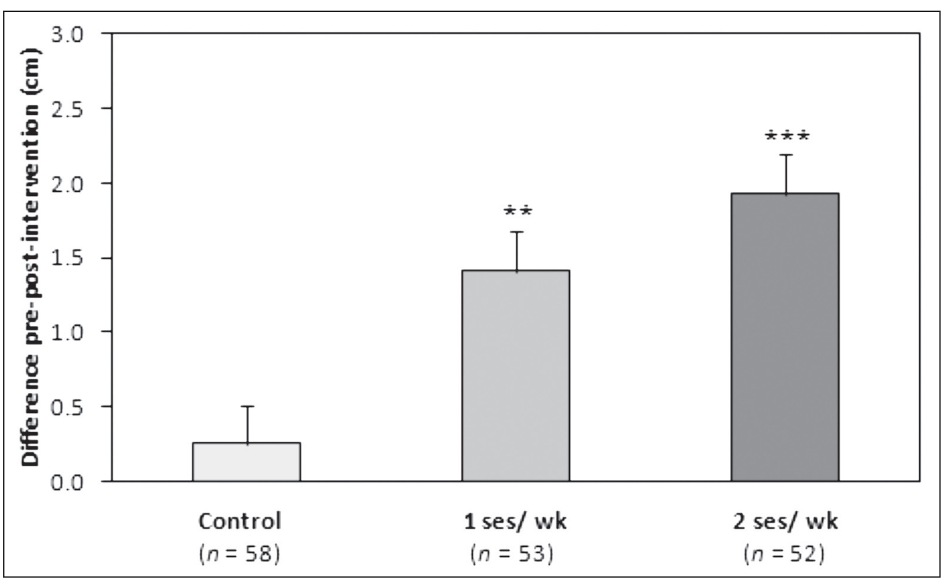


Fig. 3.—Effect of the stretching intervention program on hamstring extensibility (sit-and-reach scores, cm). Note. *Values represent the mean and error bars the standard error. Significance level from the one-way analysis of variance (*p *< 0.001) followed by the pairwise comparisons with the Bonferroni adjustment: Change statistically significant one session/ week-control group (*** p *< 0.01) and two sessions/ week-control group (**** p *< 0.001).*

a long-term program instead of a short-term one was performed as in the present study must also be highlighted.

As far as we know there is only one related study carried out on adults[18]. Marques et al.[18] compared the effects of a short-term stretching program performed one, three and five times a week on hamstring extensibility in apparently healthy adults. Similar to the present study, these authors found that a stretching intervention program carried out once a week improves adults´ stand-and-reach scores. Nevertheless, when adults stretched three times a week, their improvement was statistically higher than those who stretched once a week. Participants who stretched five times a week, however, did not find any difference with those who stretched once or three times. Moreover, statistically significant differences were not found on the angular test scores for any group. Additionally, since the previous authors did not follow a controlled design, we have to be wary of these results.

As regards to the magnitude effects of the intervention, the effect size of the present stretching program was small. However, according to Valentine and Cooper[28], we have to be aware that in educational interventions even these values of effect size could be considered as of practical relevance. In order to interpret the magnitude of an educational intervention, we must also consider the ratio cost/ benefit[29]. Therefore, if an educational intervention produces even a slight increase in a key issue of a students' health marker such as hamstring extensibility and it is not a major cost (i.e., economic, labor and effort of the teacher, etc.). it may be worth keeping that intervention[29]. In addition, the results of the minimal detectable change analyses showed that the increase in the SR scores due to the intervention program was true and reliable rather than the measurement error.

Previous studies examining the effect of a short-term stretching program (5-10 weeks) carried out twice a week obtained similar or higher effect sizes than the current study (g = 0.24-0.67)[13-15,23,30,31]. Increasing training factors such as the frequency or duration of the intervention program could have a positive outcome on the magnitude effects. Regarding the frequency, Santonja et al.[16] found that when schoolchildren performed four sessions per week instead of two, the magnitude effect doubled (g = 0.85 vs. 1.53). Although in the present study the stretching program performed twice a week was slightly greater than when it was performed once a week, that difference was trivial (Δg = 0.07). However, as mentioned above, a long-term intervention program could show a greater difference as in the Santonja et al.[16] study. In this line, as regards the duration of the program, while the median effects size of previous studies for short-term was 0.43, the magnitude was higher for the mid-term stretching programs (16 weeks) (g = 0.86)[32], and even higher for those with long-term stretching programs (whole school year, 31-32 weeks) (g = 0.94)[16,33-35].

In conclusion, to our knowledge this is the first study that compares the effects of a stretching intervention program performed once and twice a week on hamstring extensibility among schoolchildren in a PE setting. The main result of the present study suggests that a short-term stretching program performed only once a week improves hamstring extensibility in schoolchildren. Additionally, when the stretching program was performed twice a week, the improvement in students´ hamstring extensibility was not significantly statistically higher. This knowledge could help and guide teachers to design programs that guarantee a feasible and effective development of students' flexibility in the PE setting. Future research studies should compare the effect of long-term stretching programs with one and two sessions per week, as well as the research into the effectiveness of stretching intervention programs with other frequencies and volume per session is also required.

Acknowledgements

We thank Marla Trush for the English revision. Daniel Mayorga-Vega is supported by a research grant from the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

Conflict of interest

None.

References

1. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, et al. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Med Sci Sports Exerc* 2011;47(7):1334–59.
2. Kendall FP, McCreary EK, Provance PG, Rodgers MM, Romani WA. Muscles: Testing and function with posture and pain. 5th ed. Baltimore: Lippincott, Williams, & Wilkins; 2005.
3. Sato T, Ito T, Hirano T, Morita O, Kikuchi R, Endo N, et al. Low back pain in childhood and adolescence: A cross sectional study in Niigata City. *Eur Spine J* 2008;17(11):1441–7.
4. Jones MA, Stratton G, Reilly T, Unnithan V. Biological risk indicators for recurrent non-specific low back pain in adolescents. *Br J Sports Med* 2005;39(3):137–40.
5. Sjölie AN. Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scand J Med Sci Sports* 2004;14(3):168–75.
6. Mikkelsson L, Nupponen H, Kaprio J, Kautiainen H, Mikkelsson M, Kujala UM. Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *Br J Sports Med* 2006;40(2):107–13.
7. Hestbaek L, Leboeuf-Yde C, Kyvik KO, Manniche C. The course of low back pain from adolescence to adulthood. Eight year follow up of 9600 twins. *Spine* 2006;31(4):468–72.
8. Kujala UM, Taimela S, Salminen JJ, Oksanen A. Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonathletes. *Scand J Med Sci Sports* 1994;4(3):200–5.

9. Ramos Espada D, González Montesinos JL, Mora Vicente J. Evolution of the articular range in primary and secondary school. *Rev Int Med Cienc Act Fís Deporte* 2007;7(26):144–57.
10. Castro-Piñero J, Girela-Rejón MJ, González-Montesinos JL, Mora J, Conde-Caveda J, Sjöström M, et al. Percentile values for flexibility tests in youths aged 6 to 17 years: Influence of weight status. *Eur J Sport Sci* 2013;13(2):139–48.
11. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: A powerful marker of health. *Int J Obes* 2008;32(1):1–11.
12. Ministerio de Educación y Ciencia. Real Decreto 1613/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la educación secundaria obligatoria. *Boletín Oficial del Estado* 2007;5:677-773.
13. Mayorga-Vega D, Merino-Marban R, Vera-Estrada F, Viciana J. Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology* 2014;46(2):227–33.
14. Merino-Marban R, Mayorga-Vega D, Fernandez-Rodriguez E, Vera Estrada F, Viciana J. Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *Eur Phy Educ Rev* 2015;21(1):83–92.
15. Sánchez Rivas E, Mayorga-Vega D, Fernández Rodríguez E, Merino-Marban R. Effect of a hamstring stretching program during physical education lessons in primary education. *J Sport Health Res* 2014;6(2):159–68.
16. Santonja Medina F, Sainz De Baranda Andújar P, Rodríguez García PL, López Miñarro PA, Canteras Jordana M. Effects of frequency of static stretching on straight-leg raise in elementary school children. *J Sports Med Phys Fitness* 2007;47(3):304–8.
17. Viciana J, Mayorga-Vega D, Merino-Marban R. Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In: Todaro R (ed) Handbook of physical education research. Role of school programs, children´s attitudes and health implications. New York: Nova Science Publisher; 2014:237–52.
18. Marques AP, Vasconcelos AA, Cabral CMN, Sacco IC. Effect of frequency of static stretching on flexibility, hamstring tightness and electromyographic activity. *Braz J Med Biol Res* 2009;42(10):949–53.
19. Mayorga-Vega D, Viciana J, Cocca A. Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *J Hum Kinet* 2013;37:153–60.
20. López-Miñarro PA, Sainz de Baranda P, Rodríguez-Garcia PL. A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *J Sports Sci Med* 2009;8(1):116–22.
21. Mayorga-Vega D, Merino-Marban R, Garcia-Romero JC. Validity of sit-and-reach with plantar flexion test in children aged 10-12 years. *Rev Int Med Cienc Act Fís Deporte* 2015;39.
22. Mayorga-Vega D, Merino-Marban R, Viciana J. Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *J Sports Sci Med* 2014;13(1):1–14.
23. Mayorga-Vega D, Merino-Marban R, Garrido FJ, Viciana J. Comparison between warm-up and cool-down stretching programs on hamstring extensibility gains in primary schoolchildren. *Phys Act Rev* 2014;2:16–24.
24. Viciana J, Mayorga-Vega D. Innovative teaching units applied to physical education. Changing the curriculum management for authentic outcomes. *Kinesiology*. In press.
25. Hedges LV. Effect sizes in cluster-randomized designs. *J Educ Behav Stat* 2007;32(4):341–70.
26. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther* 2006;86(5):735–43.
27. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–8.
28. Valentine J, Cooper H. Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. Washington: What Works Clearing House; 2003.
29. Cooper H, Hedges LV, Valentine JC. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Sage; 2009.
30. Kamandulis S, Emeljanovas A, Skurvydas A. Stretching exercise volume for flexibility enhancement in secondary school children. *J Sports Med Phys Fitness* 2013;53(6):687–92.
31. Mayorga-Vega D, Merino-Marban R, Sánchez-Rivas E, Viciana J. Effect of a short-term static stretching training program followed by five weeks of detraining on hamstring extensibility in children aged 9-10 years. *J Physic Educ Sport* 2014;14(3):355–9.
32. Coledam DHC, Arruda GA, Ramos de Oliveira A. Chronic effect of static stretching performed during warm-up on flexibility in children. *Braz J Kinanthr Hum Perform* 2012;14(3):296–304.
33. Rodríguez PL, Santonja FM, López-Miñarro PA, Sáinz de Baranda JL. Effect of physical education stretching program on sit-and-reach score in schoolchildren. *Sci Sports* 2008;23(3-4):170–5.
34. Sainz de Baranda P. Flexibility training in physical education: Intervention program. *Cultura Cienc Deporte* 2009;4(10): 33–8.
35. Sainz de Baranda P, Rodríguez PL, Santonja FM, López PA, Andújar P, Ferrer V, et al. Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *J Hum Mov Stud* 2006;51(4):277–89.

Paper XI

EFFECT OF A ONE-SESSION-PER-WEEK PHYSICAL EDUCATION-BASED STRETCHING PROGRAM ON HAMSTRING EXTENSIBILITY IN SCHOOLCHILDREN

Mayorga-Vega, D., Merino-Marban, R., Redondo-Martín, F. J., & Viciana, J.

Kinesiology

Submitted

EFFECT OF A ONE-SESSION-PER-WEEK PHYSICAL EDUCATION-BASED STRETCHING PROGRAM ON HAMSTRING EXTENSIBILITY IN SCHOOLCHILDREN

Daniel Mayorga-Vega[1], Rafael Merino-Marban[2], Francisco Javier Redondo-Martín[3] and Jesús Viciana[1]

[1]*Department of Physical Education and Sport, University of Granada, Spain.*

[2]*Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Spain*

[3]*CEIP Manuel Laza Palacios, Junta de Andalucía, Spain*

Correspondence to:

Daniel Mayorga-Vega, Ph.D. student

FPU Research Fellow (AP2010-5905)

Department of Physical Education and Sport

University of Granada

Alfacar street, 18011, Granada, Spain

Phone number: (0034) 958246641

Fax number: (0034) 958244369

E-mail: dmayorgavega@gmail.com

Abstract:

The purpose of this study was to examine the effects of a one-session-per-week stretching program on hamstring extensibility among schoolchildren in the physical education setting. Thirty seven 9-year-old schoolchildren from two classes were clustered randomly assigned to an experimental group (n=19) or a control group (n=18). During the physical education classes, the experimental students performed a 3-minute stretching program once a week for the whole academic year (a total of 32 calendar weeks, but 28 weeks of intervention after excluding holidays). Hamstring extensibility (estimated by the classic sit-and-reach test) was assessed at the beginning (week 0), in the middle (week 18) and at the end (week 34) of the stretching intervention program. The results of the two-way analysis of variance showed that the physical education-based stretching program improved statistically significantly the students' sit-and-reach scores in the middle and at the end of the intervention (p<.01). Since in physical education many curricular contents need to be developed each academic year and the subject is also restricted by its limited curriculum time allocation, teachers could improve students´ hamstring extensibility by only a one-session-per-week stretching program. Therefore, in addition to the improvement of students´ flexibility levels, this intervention program might permit the regular development of other physical education curricular contents. This knowledge could help and guide teachers to design programs that guarantee a feasible and effective development of flexibility in the physical education setting.

Keywords: Flexibility program, frequency, classic sit-and-reach test, health-related physical fitness, primary school, physical education setting.

Introduction

Physical fitness is nowadays considered as one of the most important health markers in childhood (Ortega, Ruiz, Castillo, & Sjöström, 2008), where flexibility is an important component of health-related physical fitness (National Association for Sports and Physical Education, 2005). In this line, children with an adequate hamstring extensibility seem to have a lower risk of current low back pain (Feldman, Shrier, Rossignol, & Abenhaim, 2001; Jones, Stratton, Reilly, & Unnithan, 2005; Sjölie, 2004) and neck tension (Mikkelsson, et al., 2006), as well as a lower risk of low back pain later during adulthood (Hestbaek, Leboeuf-Yde, Kyvik, & Manniche, 2006; Kujala, Taimela, Salminen, & Oksanen, 1994).

Therefore, health promotion policies should also be designed to identify children with low health-related physical fitness levels, as well as to encourage them to achieve a health-enhancing physical activity levels (Ortega, et al., 2008). In many circumstances, schools may play an important role in this public health issue (Ortega, et al., 2008). Particularly, a shortened hamstring could be addressed proactively through the subject of physical education (PE) by systematically performing stretching exercises (Santonja, Sainz De Baranda, Rodríguez, López, & Canteras, 2007; Thacker, Gilchrist, Stroup, & Kimsey, 2004). In this line, nowadays in most of the countries PE teachers are required to achieve and maintain students´ health-enhancing flexibility levels (e.g., Ministerio de Educación y Ciencia, 2006; National Association for Sport and Physical Education, 2004).

Previous studies have found that a PE-based stretching program carried out twice a week improves hamstring extensibility in schoolchildren (e.g., Coledam, Arruda, & Ramos de Oliveira, 2012; Merino-Marban, Mayorga-Vega, Fernandez-Rodriguez, Vera Estrada, & Viciana, 2015; Sainz de Baranda, 2009). However, nowadays PE teachers must face several planning-related problems for developing students´ flexibility levels (Viciana, Mayorga-Vega, & Merino-Marban, 2014b). For instance, apart from the fact that many curricular contents must be developed each academic year, PE is usually restricted by its limited curriculum time allocation. Moreover, this restriction is especially accented when the number of PE sessions a week is limited to only two, which is the norm in most European countries (European Commission/ EACEA/ Eurydice, 2013).

Since stretching programs cannot be allocated a large part of PE time, the application of a one-session-per-week stretching program could be more suitable. Unfortunately, to our knowledge there are no studies examining the effect of a one-session-per-week stretching program in schoolchildren, as well as the number of related studies with adults is really scarce (Marques, Vasconcelos, Cabral, & Sacco, 2009). Currently there is a lack of scientific information about the effects of this kind of programs among children and, therefore, research in this area is required. Consequently, the purpose of the present study was to examine the effects of a one-session-per-week stretching program on hamstring extensibility among schoolchildren in the PE setting.

Methods

Participants

Thirty seven 9-year-old schoolchildren from two different third-grade PE classes of a public primary school participated in the present study. For practical reasons and the nature of the present study (i.e., intervention focused on natural groups in a school setting) a cluster randomized controlled design was used (Mayorga-Vega, Viciana, & Cocca, 2013; Merino-Marban, et al., 2015). Natural classes were assigned randomly to form one of the study groups: control group (CG) or experimental group (EG).

All the participants were free of orthopedic disorders such as episodes of hamstring and/ or lumbar injuries, fractures, surgery or pain in the spine or hamstring and/ or lumbar muscles over the past six months (López-Miñarro, Sainz de Baranda, & Rodríguez-Garcia, 2009). The inclusion criterion was to have an attendance rate of 90% or higher for PE classes during the intervention period. Children and their legal guardians were fully informed about all the features of the study and were required to sign an informed consent document. The study protocol was approved by the Ethical Committee of the University of Malaga.

Measures

Participants' hamstring extensibility was estimated by the classic sit-and-reach (SR) test (Mayorga-Vega, Merino-Marban, & Viciana, 2014c; Mayorga-Vega, Viciana, Cocca, & Merino-Marban, 2014d). The test was applied at the beginning (pretest, week 0), in the middle (posttest, week 18) and at the end (retest, week 34) of the stretching intervention program in order to examine possible changes produced. A week before of the baseline measure, a familiarization session was carried out for all children. The SR test was applied by the same tester and instrument, a wooden box with a ruler at the top (the score 23 cm corresponded to the tangent of the feet; accuracy 0.1 cm). Additionally, the measures were performed in an indoor sports facility under the same environmental conditions, on the same day of the week and at the same time for each participant.

Prior to the evaluation, the participants performed a standardized warm-up consisting of three minutes of jogging at low intensity followed by two static and bipodal hamstring stretching exercises (two 15-second sets of each exercise). Then, the participants were assessed with the SR test in sportswear and barefoot. Briefly, at the beginning of the test the participants stood in front of the box, sat with their hips flexed, knees extended and both hands on the top of the ruler. From this position, the participants had to bend the trunk forward slowly and progressively (no swings) in order to reach the furthest possible distance and to remain still for at least two seconds. Two trials were performed one minute apart, and the average was retained (Mayorga-Vega, Merino-Marban, & Garcia-Romero, in press).

Procedures

All the participants were enrolled in their standard PE classes. Additionally, the EG participants performed a stretching-based intervention program once a week for the whole academic year (a total of 32 calendar weeks, but 28 weeks of intervention after excluding holidays). On the other hand, the CG participants followed the same standard PE classes without performing stretching exercises. The intervention program was also conducted and supervised by the same PE teacher. Additionally, all the participants were urged to maintain their normal physical activity levels outside of the supervised setting during the research period.

During each intervention session, the EG participants performed static hamstring stretches for three minutes during the cool-down period (Mayorga-Vega, Merino-Marban, Garrido, & Viciana, 2014a). In the physical education setting, this kind of intervention has been called “intermittent teaching unit” (Viciana & Mayorga-Vega, in press). Each intervention session included three 20-second sets of three bipodal stretching exercises. Overall, four different stretching exercises were designed and alternated during the intervention program: (a) standing with feet together; (b) sitting with feet together; (c) standing with feet shoulders width apart, and (d) sitting with feet shoulders width apart (see Figure 1 in Merino-Marban, et al., 2015).

In all the stretching exercises, the children flexed forward at the hip, trying to maintain the spine in neutral position as much as possible until a gentle stretch was felt in the hamstrings. The knees were fully extended and toes pointed to the ceiling with no hip rotation. The stretched positions were held gently until the end point of the range was reached (i.e. stretch to the point of feeling the tightness of the hamstring muscles, but no pain). Once this position was achieved, the children held it for 20 seconds.

Statistical analyses

Descriptive statistics (means and standard deviations) for body mass, body height, body mass index, and SR scores were calculated. A one-way analysis of variance (ANOVA) was used to study the differences of the general characteristics and baseline SR scores between groups. Moreover, chi-squared analyses were carried out to test the ratio differences of gender and extracurricular sport practitioners between the two groups. Afterward, the effect of the intervention program on hamstring extensibility was examined using a two-way ANOVA applied over the SR scores, including *group* as an independent variable (CG, EG) and *time* as a dependent variable (pretest, posttest, retest). Subsequently, for the *post hoc* analyses, α values were corrected using the Bonferroni adjustment. Then, the Hedges' *g* effect size was used to examine the magnitude of intervention effects (Hedges, 2007). The test-retest reliability of SR scores was estimated using the intraclass correlation coefficient from the two-way ANOVA (Shrout & Fleiss, 1979), as well as the 95% confidence interval. All statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20). The statistical significance level was set at $p<.05$.

Results

All the participants completed the intervention program according to previously established attendance norms. The general characteristics of the participants studied are shown in Table 1. The one-way ANOVA results did not show statistically significant differences in body mass, body height, body mass index, and SR baseline values between EG and CG ($p>.05$). Furthermore, the chi-square analyses showed that the two groups had a balanced representation of boys and girls and extra-curricular sport practitioners and non-practitioners ($p>.05$).

Table 2 shows the effect of the stretching intervention program on SR scores. The results of the two-way ANOVA on the average obtained in the SR showed interaction effects between the *group* and *time* variables [$F(2, 70)=16.233$; $p<.001$; $\eta^2=.317$; $P=.998$]. Subsequently, for *post hoc* analysis, the ANOVA with the Bonferroni adjustment showed that the EG improved hamstring extensibility statistically significantly from pretest to posttest ($p=.009$), from posttest to retest ($p<.001$), and from pretest to retest ($p<.001$). On the other hand, for the CG no significant differences were found ($p=1.000$). The test-retest reliability for SR was .98 (.95-.99).

Discussion and conclusions

The purpose of the present study was to examine the effects of a one-session-per-week stretching program on hamstring extensibility among schoolchildren in the PE setting. Nowadays in most of the countries PE teachers are required to achieve and maintain students´ health-enhancing flexibility levels (e.g., Ministerio de Educación y Ciencia, 2006; National Association for Sport and Physical Education, 2004). However, PE teachers must face several planning-related problems (Viciana, et al., 2014b). For instance, PE is restricted by its limited curriculum time allocation. Moreover, another PE planning-related problem is the fact that there are many curricular contents that teachers must develop in each academic year (Viciana, Mayorga-Vega, & Cocca, 2014a; Viciana, et al., 2014b).

Consequently, since stretching programs cannot be allocated a large part of PE time, the application of a one-session-per-week stretching program could be more suitable. In this sense, the results of the present study showed that a PE-based stretching program performed only once a week improves hamstring extensibility in schoolchildren. Unfortunately, to our knowledge there are no studies examining the effect of a one-session-per-week stretching program in schoolchildren. Furthermore, as far as we know there is only one related study carried out with adults (Marques, et al., 2009). Similar to the present study, Marques, et al.

(2009) found that a stretching program carried out once a week improves adults´ hamstring extensibility. However, since the previous authors did not follow a controlled design, we have to be aware of these results.

Regarding the magnitude effects of the intervention, at the end of the academic year the effect size of the present study was moderate indicating that the stretching program was effective. In contrast with the current results, all the previous studies carried out twice a week for the whole school year (31-32 weeks) obtained large and very large effect sizes (g=0.85-2.06) (Rodríguez, Santonja, López-Miñarro, Sáinz de Baranda, & Yuste, 2008; Sainz de Baranda, 2009; Sainz de Baranda, et al., 2006; Santonja, et al., 2007). Apart from other training factors such as the time of stretching per session (i.e., in all the above-mentioned studies it was 5-7 min comparing to 3 min in the present study), the frequency might clearly influence the magnitude effect of the intervention (i.e., two sessions per week instead of once a week in the present study). In this line, Santonja, et al. (2007) found that when the children performed four sessions per week instead of two, the magnitude effect doubled (twice a week, g=0.85; four times a week, g=1.53; both 5 minutes per session, 31 weeks of duration).

The frequency of the stretching program could have a positive relationship with the effect size of the intervention. However, since in most European countries PE is limited to a low frequency (European Commission/ EACEA/ Eurydice, 2013), the application of stretching programs with a high frequency is not feasible in this setting. In this line, Merino-Marban, et al. (2015) indicated that the increase of active time for learning in extra-curricular periods would represent an excellent strategy for PE teachers to pursue important objectives such as the flexibility improvement. Nevertheless, since this strategy mainly depends on the students' autonomy, we have to be aware that in primary schoolchildren it could be impractical (Merino-Marban, et al., 2015). Therefore, although the frequency of the stretching programs seem to show a positive relationship with the effect size of the intervention, even a PE-based stretching program performed only once a week is effective among schoolchildren.

On the other hand, the duration of the intervention program also seems to be an important training factor. While the previous studies examining the long-term stretching programs effects (whole school year, 31-32 weeks) obtained a large and very large effect sizes (*g*=0.85-2.06) (Rodríguez, et al., 2008; Sainz de Baranda, 2009; Sainz de Baranda, et al., 2006; Santonja, et al., 2007), the magnitude was lower for the mid-term stretching programs (16 weeks) (*g*=0.85-0.88) (Coledam, et al., 2012), and even lower for those with short-term programs (8-10 weeks) (*g*=0.35-0.67) (Mayorga-Vega, et al., 2014a; Mayorga-Vega, Merino-Marban, Vera-Estrada, & Viciana, 2014b; Merino-Marban, et al., 2015; Sánchez Rivas, Mayorga-Vega, Fernández Rodríguez, & Merino-Marban, 2014). In this line, in the present study the effect size obtained in the mid-term (after 17 weeks of intervention) was half that in the whole academic year (after 32 weeks of intervention). Anyway, according to Valentine and Cooper (2003), we have to be aware that in education research even these values of effect size could be considered of practical relevance.

In conclusion, to our knowledge the present study is the first one that examines the effect of a one-session-per-week stretching program on hamstring extensibility among schoolchildren in the PE setting. The results of this study suggest that it is possible to improve students´ hamstring extensibility by only one 3-minute session per week in PE setting (i.e., only 3% of the PE time in a common situation of 50-minute sessions twice a week). Therefore, in addition to the improvement of students´ flexibility levels, this intervention program might permit the regular development of other PE curricular contents. This knowledge could help and guide teachers to design programs that guarantee a feasible and effective development of flexibility in the PE setting. Future research studies should compare the effect of PE-based stretching programs with different frequencies, as well as the effectiveness of short-term stretching interventions with only one session per week.

Acknowledgements

We thank Anna Szczesniak for the English revision. The first author is supported by a research grant from the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

References

Coledam, D.H.C., Arruda, G.A., & Ramos de Oliveira, A. (2012). Chronic effect of static stretching performed during warm-up on flexibility in children. *Revista Brasileira de Cineantropometria e Desempenho Humano*, *14*(3), 296-304.

European Commission/ EACEA/ Eurydice (2013). *Physical education and sport at school in Europe Eurydice Report*. Luxembourg: Publications Office of the European Union.

Feldman, D.E., Shrier, I., Rossignol, M., & Abenhaim, L. (2001). Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology*, *154*(1), 30-36.

Hedges, L.V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341-370.

Hestbaek, L., Leboeuf-Yde, C., Kyvik, K.O., & Manniche, C. (2006). The course of low back pain from adolescence to adulthood. Eight year follow up of 9600 twins. *Spine*, *31*(4), 468-472.

Jones, M.A., Stratton, G., Reilly, T., & Unnithan, V.B. (2005). Biological risk indicators for recurrent non-specific low back pain in adolescents. *British Journal of Sports Medicine*, *39*(3), 137-140.

Kujala, U.M., Taimela, S., Salminen, J.J., & Oksanen, A. (1994). Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonathletes. *Scandinavian Journal of Medicine and Science in Sports*, *4*(3), 200-205.

López-Miñarro, P.A., Sainz de Baranda, P., & Rodríguez-Garcia, P.L. (2009). A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science and Medicine*, *8*(1), 116-122.

Marques, A.P., Vasconcelos, A.A., Cabral, C.M.N., & Sacco, I.C. (2009). Effect of frequency of static stretching on flexibility, hamstring tightness and electromyographic activity. *Brazilian Journal of Medical and Biological Research*, *42*(10), 949-953.

Mayorga-Vega, D., Merino-Marban, R., & Garcia-Romero, J.C. (in press). Validity of sit-and-reach with plantar flexion test in children aged 10-12 years. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte*. Retrieved from http://cdeporte.rediris.es/revista/inpress/artvalidez602e.pdf on April 15th, 2015.

Mayorga-Vega, D., Merino-Marban, R., Garrido, F.J., & Viciana, J. (2014a). Comparison between warm-up and cool-down stretching programs on hamstring extensibility gains in primary schoolchildren. *Physical Activity Review*, *2*, 16-24.

Mayorga-Vega, D., Merino-Marban, R., Vera-Estrada, F., & Viciana, J. (2014b). Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology*, *46*(2), 227-233.

Mayorga-Vega, D., Merino-Marban, R., & Viciana, J. (2014c). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine*, *13*(1), 1-14.

Mayorga-Vega, D., Viciana, J., & Cocca, A. (2013). Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics*, *37*, 153-160.

Mayorga-Vega, D., Viciana, J., Cocca, A., & Merino-Marban, R. (2014d). Criterion-related validity of toe-touch test for estimating hamstring extensibility: A meta-analysis. *Journal of Human Sport and Exercise*, *9*(1), 188-200.

Merino-Marban, R., Mayorga-Vega, D., Fernandez-Rodriguez, E., Vera Estrada, F., & Viciana, J. (2015). Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *European Physical Education Review*, *21*(1), 83-92.

Mikkelsson, L., Nupponen, H., Kaprio, J., Kautiainen, H., Mikkelsspn, M., & Kujala, U. (2006). Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine*, *40*(2), 107-113.

Ministerio de Educación y Ciencia (2006). Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria. [Royal Decree 1513/2006, from the 7th December, laying down the Curriculum for Primary Education. In Spanish.] *Government Gazette*, *293*, 43075-43080.

National Association for Sport and Physical Education (2004). *Moving into the future: National standards for physical education*. 2nd ed. Reston, VA: Author.

National Association for Sports and Physical Education (2005). *Physical education for lifelong fitness: The physical best teacher's guide*. Leeds: Human Kinetics.

Ortega, F.B., Ruiz, J.R., Castillo, M.J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of* Obesity, *32*(1), 1-11.

Rodríguez, P.L., Santonja, F.M., López-Miñarro, P.A., Sáinz de Baranda, P., & Yuste, J.L. (2008). Effect of physical education stretching program on sit-and-reach score in schoolchildren. *Science and Sports*, *23*(3-4), 170-175.

Sainz de Baranda, P. (2009). Flexibility training in physical education: Intervention program. *Cultura, Ciencia y Deporte*, *4*(10), 33-38.

Sainz de Baranda, P., Rodríguez, P.L., Santonja, F.M., López, P.A., Andújar, P., Ferrer V., & Pastor, A. (2006). Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *Journal of Human Movement Studies*, *51*(4), 277-289.

Sánchez Rivas, E., Mayorga-Vega, D., Fernández Rodríguez, E., & Merino-Marban, R. (2014). Effect of a hamstring stretching program during physical education lessons in primary education. *Journal of Sport and Health Research*, *6*(2), 159-168.

Santonja, F.M., Sainz De Baranda, P., Rodríguez, P.L., López, P.A., & Canteras, M. (2007). Effects of frequency of static stretching on straight-leg raise in elementary school children. *Journal of Sports Medicine and Physical Fitness*, *47*(3), 304-308.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428.

Sjölie, A.N. (2004). Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scandinavian Journal of Medicine and Science in Sports*, *14*(3), 168-175.

Thacker, S.B., Gilchrist, J., Stroup, D.F., & Kimsey, C.D.J. (2004). The impact of stretching on sports injury risk: A systematic review of the literature. *Medicine and Science in Sports and Exercise*, *36*(3), 371-378.

Valentine, J., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington: What Works Clearing House.

Viciana, J., & Mayorga-Vega, D. (in press). Innovative teaching units applied to physical education. Changing the curriculum management for authentic outcomes. *Kinesiology*.

Viciana, J., Mayorga-Vega, D., & Cocca, A. (2014a). Model of successful learning in physical education and its maintenance. Study of the effect of intermittent reinforcement on physical fitness. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, *9*(1), 155-171.

Viciana, J., Mayorga-Vega, D., & Merino-Marban, R. (2014b). Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In R. Todaro (Ed.), *Handbook of physical education research. Role of school programs, children´s attitudes and health implications* (pp. 237-252). New York: Nova Science Publisher.

Table 1. General characteristics (mean±standard deviation/ frequency) of the participants and differences between experimental and control groups.

	Sample (N=37)	Experimental (n=19)	Control (n=18)	Differences[a]	
				F/χ^2	p
Body mass (kg)	36.45±8.03	35.43±5.89	37.53±9.87	.625	.434
Body height (m)	1.36±0.06	1.35±0.04	1.37±0.07	.734	.397
Body mass index (kg/m^2)	19.63±3.52	19.41±3.20	19.86±3.92	.142	.708
Gender (boys/girls)	18/19	8/11	10/8	.669	.413
Extra-curricular sport (yes/no)[b]	17/20	11/8	6/12	2.245	.134

Note. [a] Significance level from the one-way analysis of variance for the body mass, body height and body mass index, and from the chi squared test for the gender and extra-curricular sport ratios; [b]Children that regularly participated (yes) or not (no) at least twice per week in extra-curricular sport activities.

Table 2. Effect of the stretching-based intervention program on sit-and-reach scores (cm).

Group	Pretest (1) (M±SD)	Posttest (2) (M±SD)	Retest (3) (M±SD)	p[a]	Effect size[b]		
					1-2	2-3	1-3
Experimental (n=19)	24.0±5.5	25.5±5.8*	27.6±4.6^†	<.001	.28	.30	.60
Control (n=18)	24.2±7.2	23.9±7.7	24.0±7.7				

Note. M=mean; SD=standard deviation; [a]Significance level from the two-way analysis of variance with the post hoc analysis with Bonferroni adjustment: Change statistically significant from pretest to posttest (*p<.01), from posttest to retest (^p<.001), and from pretest to retest (†p<.001); [b] Hedges' *g* effect size.

Paper XII

EFFECT OF A SHORT-TERM PHYSICAL EDUCATION-BASED FLEXIBILITY PROGRAM ON HAMSTRING AND LUMBAR EXTENSIBILITY AND ITS POSTERIOR REDUCTION IN PRIMARY SCHOOLCHILDREN

Mayorga-Vega, D., Merino-Marban, R., Vera-Estrada, F., & Viciana, J.

Kinesiology

2014, *46*(2), 227-233.

EFFECT OF A SHORT-TERM PHYSICAL EDUCATION-BASED FLEXIBILITY PROGRAM ON HAMSTRING AND LUMBAR EXTENSIBILITY AND ITS POSTERIOR REDUCTION IN PRIMARY SCHOOLCHILDREN

Daniel Mayorga-Vega[1], Rafael Merino-Marban[2], Francisco Vera-Estrada[3] and Jesús Viciana[1]

[1]*Department of Physical Education and Sport, University of Granada, Spain*
[2]*Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Spain*
[3]*Department of Physical Education, Primary Education School of Tartessos, Junta de Andalucía, Spain*

Original scientific paper
UDC: 796.015:615.825:616.7-373.3-053.5(460)

Abstract:

The purpose of this study was to examine the effects of a short-term flexibility program on hamstring and lumbar extensibility and its posterior reduction among primary schoolchildren in a physical education (PE) setting. Forty-five 10-to-11-year-old schoolchildren from two classes were clustered randomly to an experimental group (EG) (n=22) or a control group (CG) (n=23). During the PE classes, the students in EG performed a six-minute flexibility program twice a week for eight weeks. Subsequently, these students underwent a five-week detraining period. The results of the two-way ANOVA showed that the intervention program significantly increased the students' hamstring and lumbar extensibility (pretest=15.7±7.0 cm; posttest=18.2±7.7 cm; p<.001). Although after the detraining period flexibility levels decreased statistically significantly (retest=17.1±7.9 cm; p<.001), the students from EG presented statistically higher values than in the baseline flexibility level (p=.006). For the CG no significant differences were found (pretest=13.4±8.5 cm; posttest=13.1±8.5 cm; retest=13.2±8.4 cm; p=1.000). Although children lose a significant part of the obtained flexibility gains over a five-week detraining period, they do not revert to their baseline flexibility level. Hence, the students might continue working on their flexibility within the next five weeks in order to maintain the gains obtained previously. These findings could help teachers to design programs that guarantee feasible improvement and maintenance of children's flexibility in a physical education setting.

Key words: *stretching program, detraining, elementary schoolchildren, physical education setting, classical sit-and-reach test, health-related physical fitness*

Introduction

Hamstring and lumbar extensibility is a recognized health-related physical fitness component that plays an important role in protecting the spine from possible risks and, therefore, allowing people to execute the normal daily living activities and social functioning (Roth-Isigkeit, Thyen, Stöven, Schwarzenberger, & Schmucker, 2005; Sato, et al., 2008). For instance, poor hamstring extensibility is associated with several spinal disorders such as thoracic hyperkyphosis (Fisk, Baigent, & Hill, 1984), spondylolysis (Standaert & Herring, 2000), disc herniation (Harvey & Tanner, 1991), changes in lumbopelvic rhythm (López-Miñarro & Alacid, 2009) and low back pain (Sjölie, 2004). Specifically in children, poor hamstring and lumbar extensibility has been associated with current low back pain (Feldman, Shrier, Rossignol, & Abenhaim, 2001; Jones, Stratton, Reilly, & Unnithan, 2005; Sjölie, 2004) and neck tension (Mikkelsson, et al., 2006), as well as with a higher risk of low-back pain during adulthood (Hestbaek, Leboeuf-Yde, Kyvik, & Manniche, 2006; Kujala, Taimela, Salminen, & Oksanen, 1994).

Limited hamstring and lumbar extensibility affects a large number of schoolchildren (Brodersen, Pedersen, & Reimers, 1994; Harreby, et al., 1999). For instance, in Spain about 18-38% of schoolchildren show reduced hamstring and lumbar extensibility (Ferrer, 1998; Castro-Piñero, et al., 2013). Shortened hamstring and lumbar muscles are the locomotor pathology more likely to be addressed proactively in physical education (PE) because their treatment is based on the execution of stretching exercises and postural correction (Santonja, Rodríguez, Sainz de Baranda, & López, 2004; Santonja, Sainz de Baranda, Rodríguez, López, & Canteras, 2007; Thacker, Gilchrist, Stroup, & Kimsey, 2004). Therefore, it seems that PE teachers should include stretching exercises within their classes (Kanásová, 2008; Rodríguez, Santonja, López-Miñarro, Sáinz de Baranda, & Yuste, 2008; Sainz de Baranda, et al., 2006; Santonja, et al., 2007).

Previous studies found that PE-based flexibility programs performed twice a week for 16-32 weeks improve hamstring and lumbar extensibility in primary schoolchildren (Coledam, Arruda, & Ramos de Oliveira, 2012; Rodríguez, et al., 2008; Rodríguez, et al., 1999; Sainz de Baranda, et al., 2006). However, a planning-related problem in PE is that the teachers must "deliver" a large volume of curricular contents during each academic course (Hardman, 2008; Ministerio de Educación y Ciencia, 2006). Hence, the application of a short-term flexibility development program seems to be more suitable (Viciana, Mayorga-Vega, & Cocca, 2013, 2014). Furthermore, since the academic year is frequently interrupted by several holiday periods, another problem related to the PE planning is that after a period of detraining the obtained flexibility gains are expected to decrease (Cipriano, Terry, Haines, Tabibnia, & Lyssanova, 2012; Rancour, Holmes, & Cipriani, 2009; Willy, Kyle, Moore, & Chleboun, 2001). Unfortunately, all of the previous studies about flexibility detraining were carried out with adults, and were limited and contradictory (Cipriano, et al., 2012; Rancour, et al., 2009; Willy, et al., 2001). A flexibility maintenance program should be applied in order to maintain the flexibility levels previously gained during the remainder of the academic year (Viciana, et al., 2013, 2014). However, the current scientific information about the efficacy of this kind of programs is still limited (Rancour, et al., 2001), especially in a PE setting.

Hence, PE teachers include stretching exercises in their classes only for a few weeks, without knowing how long the effects of these exercises will last. Unfortunately, to our knowledge there are no studies examining the effect of a short-term flexibility program and its posterior detraining among schoolchildren. Consequently, the purposes of this study were: (a) to examine the effects of a short-term PE-based flexibility program on hamstring and lumbar extensibility in schoolchildren aged 10-11 years; and (b) to evaluate the effects of a five-week period of flexibility detraining on hamstring and lumbar extensibility in schoolchildren aged 10-11 years. It was hypothesized that an eight-week PE-based flexibility program would develop hamstring and lumbar extensibility in schoolchildren, as well as that after a five-week period of detraining schoolchildren's flexibility levels would, at least partially, be lost. Regrettably, because of the lack of previous studies regarding the flexibility detraining in children together with the limited related information among adults, a more certain hypothesis cannot be postulated.

Methods

Participants

A sample of 45 schoolchildren, 26 boys and 19 girls, aged 10-11 years, from two different sixth grade PE classes of a public primary school participated in the present study. For practical reasons and the nature of the present study (the intervention was focused on natural groups in a school setting) a cluster randomized controlled trial was used (Viciana, et al., 2013). Natural classes were assigned randomly to form one of the following study groups: control group (CG) or experimental group (EG).

All the participants were free of orthopedic disorders such as episodes of hamstring and/or lumbar injuries, fractures, surgery or pain in the spine or hamstring/lumbar muscles over the past six months (López-Miñarro, Sainz de Baranda, & Rodríguez-Garcia, 2009). The inclusion criterion was to have an attendance rate of 90% or higher for PE classes during the intervention period. Children and their parents or legal guardians were fully informed about all the features of the study, and were required to sign an informed-consent document. The Ethical Committee of the University of Malaga approved the study protocol.

Measures

Hamstring and lumbar extensibility was estimated using the classical sit-and-reach (SR) test (Mayorga-Vega, Merino-Marban, & Viciana, 2014). The SR test was applied at the beginning and at the end of the flexibility intervention program (pretest and, after eight weeks, posttest, respectively) in order to examine the possible changes produced. Then, after five weeks of flexibility detraining a reassessment was performed in order to observe the levels of retention (retest).

Hamstring and lumbar extensibility was assessed by the same tester, instruments and under the same conditions. The SR test was administered using a wooden box with a ruler on the top (the score of 15 cm corresponded to the tangent of the feet; accuracy 0.1 cm). The measures were taken in an indoor sports facility under the same environmental conditions, on the same day of the week and at the same time for each student. No warm-up exercises were performed prior to the flexibility measurements.

At the beginning of the test the children stood in front of the box, sat with their hips flexed, knees extended and both hands on the top of the ruler. The feet were placed to the width of the hips and ankles at 90º. The knees were fixed in extension with the help of the tester. The hands with the fingers extended were placed parallel. From this position, the children had to bend the trunk forward slowly and progressively (no swings) in order to reach the furthest possible distance and to remain still for at least two seconds. The average of two attempts was retained (Mayorga-Vega, Merino-Marban, & Garcia-Romero, in press).

Procedures

The EG participants performed a flexibility intervention program during their regular PE classes. The flexibility program was conducted and supervised by the same PE teacher in both groups. Firstly, the EG students performed a flexibility development intervention program twice a week on non-consecutive days for eight weeks. Subsequently, coinciding with the Christmas holidays, the EG participants underwent a five-week period of flexibility detraining. Similarly to previous studies carried out in a PE setting (Coledam, et al., 2012; Rodríguez, et al., 2008; Sainz de Baranda, et al., 2006; Santonja, et al., 2007), the EG students performed hamstring/lumbar stretches using the static technique for six minutes during the cool-down period of their regular PE classes.

Each intervention session included three 20-second sets of five stretching exercises. Six different stretching exercises were designed and alternated during the intervention program (Figure 1). Four bidopal exercises and one unipodal exercise were performed in each session. In all the stretching exercises, the children flexed forward at the hip, maintaining the spine in a neutral position until a gentle stretch was felt in the hamstrings. The knees were fully extended and toes pointed to the ceiling with no hip rotation. The stretched positions were held gently until the end point of the range was reached (i.e. stretch to the point of feeling the tightness of the hamstring muscles, but no pain).

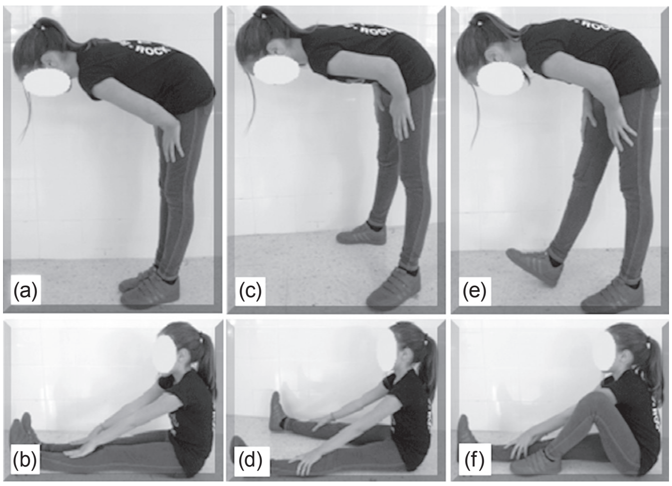


Figure 1. The six stretching exercises performed during the intervention program: (a) standing with feet together; (b) sitting with feet together; (c) standing with feet shoulder-width apart; (d) sitting with feet shoulder-width apart; (e) standing with only one leg extended, and (f) sitting with only one leg extended.

All the participants were urged to maintain their normal levels of physical activity outside the supervised setting during the research period. Twelve students in the EG (55%) and seven students in the CG (30%) regularly participated (at least twice per week) in organized extra-curricular sport programs. During the flexibility program period all the students participated in their standard PE classes. However, the CG participants followed the standard PE program without performing stretching exercises and were not aware of the purpose of the study.

Statistical analyses

Descriptive statistics (means and standard deviations) for age, body mass, body height, body mass index, and SR scores were calculated. A one-way analysis of variance (ANOVA) was used to study the differences in body mass, body height, body mass index, and baseline SR scores between EG and CG. Additionally, chi-squared analyses were carried out to test the ratio differences of gender and extra-curricular sport practitioners between the two groups. Subsequently, the effect of the eight-week flexibility intervention program followed by five weeks of detraining on hamstring and lumbar extensibility was examined using a two-way ANOVA applied over the SR scores, including *group* as an independent variable (CG, EG) and *time* as a dependent variable (pretest, posttest, retest). For the *post-hoc* analyses, α values were corrected using the Bonferroni adjustment. Moreover, the Hedges' g effect size was used to examine the magnitude of treatment effects (Hedges, 2007). The test-retest reliability of the SR test was estimated using the intra-class correlation coefficient from one-way ANOVA ($ICC_{1,2}$) (Shrout & Fleiss, 1979), as well as the 95% interval of confidence. All statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20). The statistical significance level was set at $p<.05$.

Results

All the participants completed the intervention program according to previously established attendance norms, the EG students obtaining an average attendance over 95%. The general characteristics of the participants studied are shown in Table 1. The one-way ANOVA results did not show statistically significant differences in body mass, body height, body mass index, and SR baseline values between EG and CG ($p>.05$). Additionally, the chi-square analyses showed that the two groups had a balanced representation of boys and girls and extra-curricular sport practitioners and non-practitioners ($p>.05$).

Table 2 shows the effect of the flexibility intervention program on hamstring and lumbar extensibility. The results of the two-

Table 1. General characteristics (mean±standard deviation/frequency) of the participants and differences between experimental and control groups

	Sample (N=45)	Experimental (n=22)	Control (n=23)	Differences[a]	
				F	p
Age (year)	10.9±0.3	10.9±0.3	10.9±0.3	-	-
Body mass (kg)	40.4±6.7	41.3±8.3	39.6±4.8	.685	.412
Body height (cm)	145.9±6.5	147.1±6.5	144.8±6.4	1.489	.229
Body mass index (kg/m^2)	19.0±2.9	19.0±3.4	18.9±2.4	.009	.924
Gender (boys/girls)	26/19	14/8	12/11	.606	.436
Extra-curricular sport (yes/no)[b]	19/26	12/10	7/16	2.680	.102

Note. [a]Significance level from the analysis of variance for the body mass, body height and body mass index, and from the chi-square test for the gender and extra-curricular sport ratios. [b]Children that regularly participated (yes) or not (no) at least twice per week in organized extra-curricular sport activities.

Table 2. Effect of the flexibility intervention program on the classical sit-and-reach scores (cm)

Group	Pretest (1) (M±SD)	Posttest (2) (M±SD)	Retest (3) (M±SD)	p[a]	Effect size[b]		
					1-2	2-3	1-3
Experimental (n=22)	15.7±7.0	18.2±7.7***	17.1±7.9***††	<.001	.35	-.15	.20
Control (n=23)	13.4±8.5	13.1±8.5	13.2±8.4				

Note. M=mean; SD=standard deviation; [a] Significance level from two-way analysis of variance with the *post-hoc* analysis with Bonferroni adjustment: change statistically significant from pretest to posttest (***p<.001), from posttest to retest (***p<.001), and from pretest to retest (††p<.01). [b] Hedges' *g* effect size.

way ANOVA on the average obtained in the SR showed interaction effects between the *group* and *time* variables [F(2, 86)=15.657; p<.001; η^2_p=.267; P=.997]. Subsequently, for *post-hoc* analyses, the ANOVA with the Bonferroni adjustment showed that EG improved flexibility statistically significantly from pretest to posttest (p<.001). In addition, although the flexibility levels from posttest to retest decreased statistically significantly (p<.001), EG presented statistically higher values in retest than in the baseline flexibility levels (p=.006). For CG no significant differences were found (p=1.000). The test-retest reliability for SR was .997 (.994-.999).

Discussion and conclusions

The first purpose of the present study was to examine the effects of a short-term PE-based flexibility program on hamstring and lumbar extensibility in primary schoolchildren. The results of this study showed that a six-minute PE-based flexibility development program, performed twice a week for only eight weeks, improved hamstring and lumbar extensibility in schoolchildren. In this line, previous studies in which primary schoolchildren performed a PE-based flexibility program found a significant improvement in hamstring and lumbar extensibility (Coledam, et al., 2012; Rodríguez, et al., 2008; Rodríguez, et al., 1999; Sainz de Baranda, et al., 2006).

However, despite the fact that in a PE setting the application of short flexibility development programs seems to be more feasible, in all the above mentioned studies the EG students carried out the flexibility intervention programs with a considerably longer duration, lasting from 16 (Coledam, et al., 2012) to 31-32 weeks (Rodríguez, et al., 2008; Rodríguez, et al., 1999; Sainz de Baranda, et al., 2006). On the other hand, similarly to the present research, in previous short-term stretching programs carried out with adults, improvements in flexibility were found (Cipriano, et al., 2012; Rancour, et al., 2009; Willy, et al., 2001). Although in these studies the flexibility gains were found after only 4-6 weeks, the participants performed stretching exercises from 3-4 to 14 times a week. Nevertheless, as in many other countries (European Commission/ EACEA/ Eurydice, 2013), PE in Spain is limited to two sessions a week and, therefore, stretching programs with higher frequencies are not feasible in this setting.

The second purpose of the present study was to evaluate the effect of a five-week period of flexibility detraining on hamstring and lumbar extensibility in schoolchildren. In PE planning another common limitation is the fact that an academic year is frequently interrupted by several holiday periods. Therefore, although after a period of detraining the flexibility gains obtained are expected to decrease (Cipriano, et al., 2012; Rancour, et al., 2009; Willy, et al., 2001), PE teachers include stretching exercises in their classes only for a few weeks, and then they cease doing them because of the holiday periods or the necessity to teach other curricular contents. Unfortunately, the current scientific information about the flexibility detraining is limited and contradictory (Cipriano, et al., 2012; Rancour, et al., 2009; Willy, et al., 2001). Similarly to the present study, Cipriano et al. (2012) and Rancour et al. (2009) found that after four weeks of flexibility detraining adults retained significant gains. However, on the contrary, Willy et al. (2001) observed that adults' flexibility levels decreased to baseline after four weeks of detraining. Additionally, to our knowledge there are no previous studies examining the flexibility loss effects during detraining periods among schoolchildren.

Although most previous studies examined the effects of a flexibility program over the whole academic year (Rodríguez, et al., 2008; Rodríguez, et al., 1999; Sainz de Baranda, et al., 2006), they only examined the global effects ignoring the changes in flexibility during relatively long detraining periods such as the Christmas holidays. Therefore, one of the most important outcomes of the current study was to show that, although in a five-week detraining period children lost a significant portion of their flexibility gains previously obtained, they did not revert to their baseline flexibility level. Hence, as the effect of an eight-week flexibility program performed for six minutes twice a week is not completely worn off, PE teachers should continue to develop students' flexibility within the next five weeks in order to maintain the gains obtained in the previous semester.

Unfortunately, scientific information about the efficacy of flexibility maintenance programs among schoolchildren has not been found and the studies among adults are limited. In this line, Rancour et al. (2009) found out that, after a daily development stretching program, with 2-3 sessions a week, the adults maintained the flexibility levels previously gained. Regrettably, as it was pointed out before, since in the most countries PE is performed only twice a week, the application of this program is not suitable. In order to apply this program in a PE setting, for instance, the efficacy of a maintenance program of one session a week or the half volume of stretching in each session should be tested instead. Additionally, Rancour et al. (2009) applied the maintenance program just after the development program, that is, without a period of inactivity between the development and maintenance. However, in most countries, the efficacy of a maintenance program should be examined after a period of detraining because it is the most common situation in normal PE planning (due to the typical alternation of holidays, academic periods and the need to teach other curricular contents in the PE classes) (Viciana, et al., 2013, 2014). Regrettably, previous studies evaluating the effects of the flexibility maintenance program in a PE setting have not been found.

Regarding the physiological explanation of the increase in hamstring and lumbar extensibility observed after the development stretching program, a few theories have been proposed (Weppler & Magnusson, 2010). Traditionally, most of these theories suggest that increases in muscle extensibility observed after a flexibility program might be due to a mechanical increase in length of the stretched muscle (e.g. due to the viscoelastic deformation, plastic deformation, increased sarcomeres in series, and/or neuromuscular relaxation). However, a new theory has been proposed recently suggesting that increases in muscle extensibility could be due to a modification of sensation. In this line, increases in muscle extensibility observed especially after a short-term flexibility program could be, at least predominantly, due to modifications in individuals' sensation. Likewise, all these theories may explain the reduction in hamstring and lumbar extensibility during the detraining stretching period. A more extensive explanation of physiological mechanisms of improvements in muscle extensibility due to a stretching program can be found in several published review articles (e.g. Gajdosik, 2001; Magnusson, 1998; Weppler & Magnusson, 2010).

In conclusion, to our knowledge this is the first study that examines the effect of a short-term flexibility developmental program lasting eight weeks followed by five weeks of detraining on hamstring and lumbar extensibility among schoolchildren. The results of the current study suggest that it is possible to improve students' hamstring and lumbar extensibility performing a PE-based flexibility program for only eight weeks. Additionally, another contribution of the present study is it has demonstrated that although children lose a significant portion of the flexibility gains previously obtained in a 5-wk detraining period, they do not revert to their baseline flexibility level. Hence, PE teachers should continue working on students' flexibility within the next five weeks in order to maintain their gains obtained previously.

For all the previously mentioned reasons, it would be beneficial for PE teachers to know the minimum duration of a flexibility program that would provide authentic outcomes, then how long it takes to lose the improvements achieved after such a development program, and how a maintenance flexibility program for students should be applied (Viciana, et al., 2014). Consequently, future research interventions should examine the effect of different detraining periods among schoolchildren, as well as the application of different maintenance training programs in order to maintain the flexibility gains obtained previously (Viciana, et al., 2013, 2014). Additionally, since PE teachers must also "deliver" a large volume of curricular contents so that flexibility cannot be allocated a large part of PE time, the effectiveness of flexibility programs consisting of sessions with shorter duration should also be examined. This knowledge could help PE teachers to design programs that guarantee feasible development and maintenance of flexibility in a PE setting.

References

Brodersen, A., Pedersen, B., & Reimers, J. (1994). Incidence of complaints about heel, knee and back related discomfort among Danish children, possible relation to short muscles. *Ugeskrift fot Laeger*, *156*(15), 2243-2245.

Castro-Piñero, J., Girela-Rejón, M.J., González-Montesinos, J.L., Mora, J., Conde-Caveda, J., Sjöström, M., & Ruiz, J.R. (2013). Percentile values for flexibility tests in youths aged 6 to 17 years: Influence of weight status. *European Journal of Sport Science*, *13*(2), 139-148.

Cipriano, D.J., Terry, M.E., Haines, M.A., Tabibnia, A.P., & Lyssanova, O. (2012). Effect of stretch frequency and sex on the rate of gain and rate of loss in muscle flexibility during a hamstring-stretching program: A randomized single-blind longitudinal study. *Journal of Strength & Conditioning Research*, *26*(8), 2119-2129.

Coledam, D.H.C., Arruda, G.A., & Ramos de Oliveira, A. (2012). Chronic effect of static stretching performed during warm-up on flexibility in children. *Revista Brasileira de Cineantropometria e Desempenho Humano*, *14*(3), 296-304.

European Commission / EACEA/ Eurydice. (2013). *Physical education and sport at school in Europe Eurydice Report*. Luxembourg: Publications Office of the European Union.

Feldman, D.E., Shrier, I., Rossignol, M., & Abenhaim, L. (2001). Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology*, *154*(1), 30-36.

Ferrer, V. (1998). *Repercusiones de la cortedad isquiosural sobre la pelvis y el raquis lumbar.* [Impact of the shortened hamstring on the pelvis and lumbar spine. In Spanish.] (Unpublished doctoral dissertation, University of Murcia, Spain).

Fisk, J.W., Baigent, M.L., & Hill, P.D. (1984). Scheuermann's disease. Clinical and radiological survey of 17 and 18 year olds. *American Journal of Physical Medicine*, *63*(1), 18-30.

Gajdosik, R.L. (2001). Passive extensibility of skeletal muscle: Review of the literature with clinical implications. *Clinical Biomechanics*, *16*(2), 87-101.

Hardman, K. (2008). Physical education in schools: A global perspective. *Kinesiology*, *40*(1), 5-28.

Harreby, M., Neergaard, K., Jessen, T., Larsen, E., Storr-Paulsen, A., Lindahl, A., … Laegaard, E. (1999). Risk factors for low back pain in a cohort of 1389 Danish school children: An epidemiologic study. *European Spine Journal*, *8*(6), 444-450.

Harvey, J., & Tanner, S. (1991). Low back pain in young athletes: A practical approach. *Sports Medicine*, *12*(6), 394-406.

Hedges, L.V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341-370.

Hestbaek, L., Leboeuf-Yde, C., Kyvik, K.O., & Manniche, C. (2006). The course of low back pain from adolescence to adulthood. Eight year follow up of 9600 twins. *Spine*, *31*(4), 468-472.

Jones, M.A., Stratton, G., Reilly, T., & Unnithan, V.B. (2005). Biological risk indicators for recurrent non-specific low back pain in adolescents. *British Journal of Sports Medicine*, *39*(3), 137-140.

Kanásová, J. (2008). Reducing shortened muscles in 10-12-year-old boys through a physical exercise program. *Medicina Sportiva*, *12*(4), 115-123.

Kujala, U.M., Taimela, S., Salminen, J.J., & Oksanen, A. (1994). Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonathletes. *Scandinavian Journal of Medicine and Science in Sports*, *4*(3), 200-205.

López-Miñarro, P.A., & Alacid, F. (2009). Influence of hamstring muscle extensibility on spinal curvatures in young athletes. *Science & Sports*, *25*(4), 188-193.

López-Miñarro, P.A., Sainz de Baranda, P., & Rodríguez-Garcia, P.L. (2009). A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science & Medicine*, *8*(1), 116-122.

Magnusson, S.P. (1998). Passive properties of human skeletal muscle during stretch maneuvers. *Scandinavian Journal of Medicine & Science in Sports*, *8*(2), 65-77.

Mayorga-Vega, D., Merino-Marban, R., & Garcia-Romero, J.C. (in press). Validity of sit-and-reach with plantar flexion test in children aged 10-12 years. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte*. Retrieved from http://cdeporte.rediris.es/revista/inpress/artvalidez602e.pdf on December 15th, 2013.

Mayorga-Vega, D., Merino-Marban, R., & Viciana, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine*, *13*(1), 1-14.

Mikkelsson, L.O., Nupponen, H., Kaprio, J., Kautiainen, H., Mikkelsspn, M., & Kujala, U. (2006). Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine*, *40*(2), 107-113.

Ministerio de Educación y Ciencia. (2006). Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria. [Royal Decree 1513/2006, from the 7th December, laying down the Curriculum for Primary Education. In Spanish.] *Government Gazette*, *293*, 43075-43080.

Rancour, J., Holmes, C.F., & Cipriani, D.J. (2009). The effects of intermittent stretching following a 4-week static stretching protocol: A randomized trial. *Journal of Strength & Conditioning Research*, *23*(8), 2217-2222.

Rodríguez, P.L., Santonja, F., Canteras, M., Delgado, M.., Fernández, J., & Balsalobre, J. (1999). Mejora de la extensibilidad isquiosural tras un programa escolar de estiramientos. [Improvement of hamstring extensibility after school program of stretching. In Spanish.] *Selección*, *8*(4), 157-164.

Rodríguez, P.L., Santonja, F.M., López-Miñarro, P.A., Sáinz de Baranda, P., & Yuste, J.L. (2008). Effect of physical education stretching program on sit-and-reach score in schoolchildren. *Science & Sports*, *23*(3), 170-175.

Roth-Isigkeit, A., Thyen U., Stöven, H., Schwarzenberger, J., & Schmucker, P. (2005). Pain among children and adolescents: Restrictions in daily living and triggering factors. *Pediatrics*, *115*(2), e152-e162.

Sainz de Baranda, P., Rodríguez, P.L, Santonja, F.M., López, P.A., Andújar, P., Ferrer V., & Pastor, A. (2006). Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *Journal of Human Movement Studies*, *51*(4), 277-289.

Santonja, F., Rodríguez, P.L., Sainz de Baranda, P., & López, P.A. (2004). Papel del profesor de educación física ante las desalineaciones de la columna vertebral. [Role of physical education teacher to the misalignment of the spine. In Spanish.] *Selección*, *13*(1), 5-17.

Santonja, F.M., Sainz de Baranda, P., Rodríguez, P.L., López, P.A., & Canteras, M. (2007). Effects of frequency of static stretching on straight-leg raise in elementary school children. *Journal of Sports Medicine and Physical Fitness*, *47*(3), 304-308.

Sato, T., Ito, T., Hirano, T., Morita, O., Kikuchi, R., Endo, N., & Tanabe, N. (2008). Low back pain in childhood and adolescence: A cross sectional study in Niigata City. *European Spine Journal*, *17*(11), 1441-1447.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428.

Sjölie, A.N. (2004). Low back pain in adolescents is associated with poor hip mobility and high body mass index. *Scandinavian Journal of Medicine & Science in Sports*, *14*(3), 168-175.

Standaert, C.J., & Herring, S.A. (2000). Spondylolysis: A critical review. *British Journal of Sports Medicine*, *34*(6), 415-422.

Thacker, S.B., Gilchrist, J., Stroup, D.F., & Kimsey, C.D.J. (2004). The impact of stretching on sports injury risk: A systematic review of the literature. *Medicine and Science in Sports and Exercise*, *36*(3), 371-378.

Viciana, J., Mayorga-Vega, D., & Cocca, A. (2013). Effects of a maintenance resistance training program on muscular strength in schoolchildren. *Kinesiology*, *45*(1), 82-91.

Viciana, J., Mayorga-Vega, D., & Cocca, A. (2014). Modelo de aprendizaje exitoso en educación física y su mantenimiento. Estudio del efecto del refuerzo intermitente sobre la condición física. [Model of successful learning in physical education and its maintenance. Study of the effect of intermittent reinforcement on physical fitness. In Spanish.] *Revista Iberoamericana de Psicología del Ejercicio y Deporte*, *9*(1), 155-171.

Weppler, C.H., & Magnusson, S.P. (2010). Increasing muscle extensibility: A matter of increasing length or modifying sensation? *Physical Therapy*, *90*(3), 438-449.

Willy, R.W., Kyle, B.A., Moore, S.A., & Chleboun, G.S. (2001). Effect of cessation and resumption of static hamstring muscle stretching on joint range of motion. *Journal of Orthopaedic and Sports Physical Therapy*, *31*(3), 138-144.

Submitted: January 28, 2014
Accepted: July 25, 2014

Correspondence to:
Daniel Mayorga-Vega, Ph.D. student
FPU Research Fellow (AP2010-5905)
University of Granada
Alfacar Street, 18011, Granada, Spain
Phone: (0034) 958246641
Fax number: (0034) 958244369
E-mail: dmayorgavega@gmail.com

Acknowledgement

We thank Anna Szczesniak for the English revision. The first author is supported by a research grant from the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

Paper XIII

EFFECT OF A PHYSICAL EDUCATION-BASED STRETCHING PROGRAMME ON SIT-AND-REACH SCORE AND ITS POSTERIOR REDUCTION IN ELEMENTARY SCHOOLCHILDREN

Merino-Marban, R., Mayorga-Vega, D., Fernandez-Rodriguez, E., Vera, F., & Viciana, J.

European Physical Education Review

2015, *21*(1), 83-92.

Article

European Physical Education Review
2015, Vol. 21(1) 83–92

DOI: 10.1177/1356336X14550942
epe.sagepub.com

Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren

Rafael Merino-Marban
University of Málaga, Spain

Daniel Mayorga-Vega
University of Granada, Spain

Emilio Fernandez-Rodriguez
University of Málaga, Spain

Francisco Vera Estrada
Elementary Education School Tartessos, Spain

Jesús Viciana
University of Granada, Spain

Abstract

The purpose of this study was to examine the effects of a 1-minute stretching programme and 5 weeks of detraining on sit-and-reach score among schoolchildren aged 5–6 years in a physical education setting. Forty-five schoolchildren 5–6 years old from two classes were clustered randomly assigned to an experimental group ($n = 23$) or a control group ($n = 22$). During the physical education classes, the students of the experimental group performed a 1-minute stretching programme twice a week for 8 weeks. Subsequently, these participants underwent a 5-week detraining period. The classic sit-and-reach test was performed at the beginning and at the end of the development programme, as well as at the end of the detraining period. The results of the two-way ANOVA showed that the intervention programme increased significantly the students' sit-and-reach scores ($p < 0.001$). However, after 5 weeks of detraining, children's flexibility reverted back to the baseline levels ($p > 0.05$). Although an only 1-minute stretching programme

Corresponding author:
Rafael Merino-Marban, Facultad de Ciencias de la Educación, Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain.
Email: rmerino@uma.es

seems to develop the schoolchildren's flexibility, after the 5-week detraining period students' score reverts back to its initial level. This knowledge could help physical education teachers to design programmes that permit students to increase and maintain flexibility levels along the entire academic year.

Keywords

Flexibility programme, detraining, primary education, physical education setting, sit-and-reach test, physical fitness

Introduction

Flexibility is a well-recognized component of health-related physical fitness (National Association for Sports and Physical Education, 2005). For instance, the lack of hamstring muscles extensibility conditions a decrease in pelvic mobility (Kendall et al., 2005; López-Miñarro et al., 2012; Muyor et al., 2012). Therefore, when individuals with low hamstring extensibility perform a maximal trunk flexion with knees extended an increase in spinal flexion and posterior pelvic tilt occur (López-Miñarro et al., 2012; Muyor et al., 2012). This invariably leads to the biomechanical changes in the pressure distribution in the spine and consequent spinal disorders (Carregaro and Coury, 2009; Kang et al., 2013). Particularly among young people, poor hamstring extensibility seems to contribute to the increase in the risk of low back pain (Feldman et al., 2001; Jones et al., 2005; Sjölie, 2004) and neck tension (Mikkelsson et al., 2006).

Nowadays it is well known that hamstring extensibility decreases significantly during school age (Chiodera et al., 2008; Kanásová, 2008; Rodríguez et al., 2008). However, Ramos Espada et al. (2007a, 2007b) found that hamstring muscles shorten at an early age, and so they concluded that flexibility exercises should be performed even at these ages. Shortened hamstring is the locomotor pathology more likely to be addressed proactively in physical education (PE) because of its high prevalence and because its treatment is based on the realization of stretches and postural correction (Santonja et al., 2004, 2007; Thacker et al., 2004). Therefore, PE teachers should include stretching exercises within their classes (Kanásová, 2008; Peiró Velert and Devis Devís, 1995; Rodríguez et al., 2008; Sainz de Baranda et al., 2006; Santonja et al., 2007).

Previous studies found that PE-based stretching programmes carried out for 5–7 minutes improve hamstring extensibility in elementary school students (Coledam et al., 2012; Rodríguez et al., 1999, 2008; Sainz de Baranda et al., 2006; Santonja et al., 2007). However, as in many other countries, Spanish PE teachers must also 'deliver' a large volume of curricular contents during each academic course (Ministerio de Educación y Ciencia, 2006). Therefore, since stretching programmes cannot be allocated a large part of PE time, the application of shorter stretching programmes could be more suitable. Unfortunately, to our knowledge there are no studies examining the effectiveness of few-minute stretching programmes in the PE setting.

In the PE setting another problem related to flexibility is that it is expected decrease after a period of detraining (Rancour et al., 2009; Willy et al., 2001). However, many PE teachers conceive planning as 'watertight drawers' that they have to fill with curricular contents (Siedentop and Tanehill, 2000). Furthermore, another related limitation is due to the fact that the academic year is frequently interrupted by several holiday periods. Therefore, PE teachers carry out stretching exercises in their classes only for a few weeks, and when they cease doing them, they do not know how long the effect will last. Currently, as there is a lack of scientific information about stretching detraining effects among children, research in this area is required.

Consequently, the purposes of this study were: (a) to evaluate the effects of a 1-minute PE-based stretching programme on sit-and-reach (SR) score in schoolchildren aged 5–6 years; and (b) to examine the effects of a 5-week period of flexibility detraining on SR score in schoolchildren aged 5–6 years.

Material and methods

Participants

Forty-five apparently healthy children 5–6 years old from two different first-grade PE classes of a public primary school participated in this study (Table 1). For practical reasons and the nature of the present study (the intervention was focused on natural groups in a school context) a cluster randomized controlled trial was used (Mayorga-Vega et al., 2013b). Natural classes were assigned randomly to form one of the study groups: control group (CG) or experimental group (EG).

All the participants were free of orthopaedic disorders such as episodes of hamstring injuries, fractures, surgery or pain in the spine or hamstring muscles over the past 6 months (López-Miñarro et al., 2009). The inclusion criterion was to have an attendance rate of 90% or higher for PE classes during the intervention period. Children and their parents or legal guardians were fully informed about all the features of the study, and were required to sign an informed-consent document. The Ethical Committee of the University of Malaga approved the study protocol.

Measures

Participants' flexibility was assessed by the classic SR test (Mayorga-Vega et al., 2014b). The SR test was used before and after the stretching intervention programme (pretest and posttest, respectively) in order to examine the possible changes produced. Subsequently, after 5 weeks of flexibility detraining, a reassessment was performed in order to observe the levels of retention (retest). The SR test was applied by the same tester using the same equipment. The measures were performed in an indoor sports facility under the same environmental conditions, on the same day of the week and at the same time for each student. No warming up exercises were performed prior to the flexibility measurements.

Table 1. General characteristics (mean ± standard deviation) of the participants.

	All ($n = 45$)	Experimental ($n = 23$)	Control ($n = 22$)
Age (year)	5.91 ± 0.29	5.91 ± 0.29	5.91 ± 0.29
Body mass (kg)	24.87 ± 6.16	24.70 ± 6.13	25.05 ± 6.33
Height (m)	1.16 ± 0.07	1.16 ± 0.07	1.16 ± 0.07
BMI (kg/m^2)	18.26 ± 3.00	18.08 ± 2.94	18.44 ± 3.13
Gender (boys/girls)	26/19	12/11	14/8

Note: BMI: Body mass index.

The SR test was administered using a wooden box with a ruler at the top (the score 15 cm corresponded to the tangent of the feet; accuracy 1 cm). At the beginning of the test, children stood in front of the box, sat with their hips flexed, knees extended and both hands on the top of the ruler. The feet were placed to the width of the hips and ankles at 90°. The knees were fixed in extension with the help of the tester. The hands with the fingers extended were placed parallel. From this position, the children had to bend the trunk forward slowly and progressively (no rebounds) in order to reach the furthest possible distance and to remain still for at least 2 seconds. Two trials were performed 1 minute apart, and the average was retained (Mayorga-Vega et al., 2013a).

Procedures

A stretching intervention programme was applied to the EG during the PE classes. Firstly, the EG participants performed a development stretching intervention programme twice a week on non-consecutive days for 8 weeks. Subsequently, coinciding with the Christmas holiday, the EG participants underwent a 5-week period of flexibility detraining. The stretching programme was conducted and supervised by the same PE teacher of the group. Based on previous studies carried out in the PE setting (Sainz de Baranda et al., 2006; Santonja et al., 2007), the EG students performed hamstring stretches utilizing the static technique for 1 minute during the cool-down period (Mayorga-Vega et al., 2014a).

Four different stretching exercises (standing with feet together, sitting with feet together, standing with feet shoulders width apart, and sitting with feet shoulders width apart) were designed and alternated during the intervention programme (Figure 1). Standing and sitting stretching exercises were alternated between sessions. Each intervention session included three sets of one stretching exercise. For all the stretching exercises, the children flexed forward at the hip, trying to maintain the spine in neutral position as much as possible until a gentle stretch was felt in the hamstrings. The knees were fully extended and toes pointed to the ceiling with no hip rotation. The stretched positions were held gently until the end point of the range was reached (i.e. stretch to

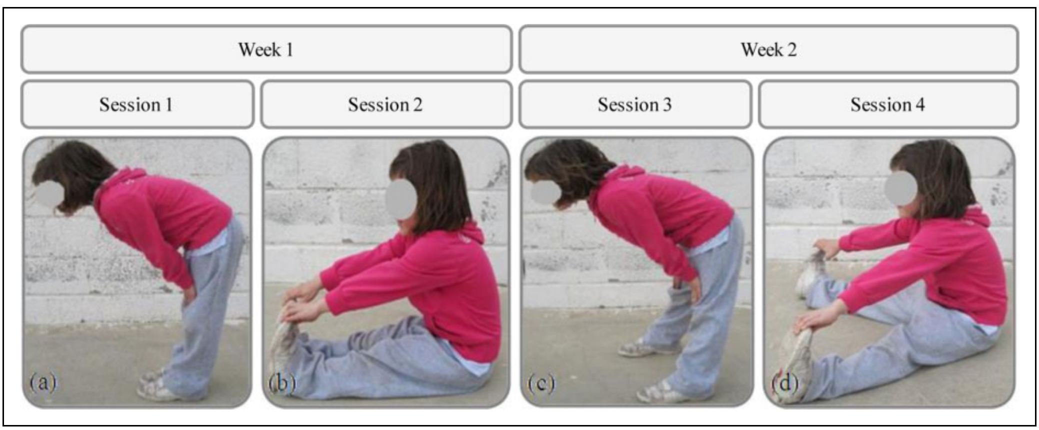


Figure 1. The four stretching exercises performed during the intervention programme: (a) standing with feet together; (b) sitting with feet together; (c) standing with feet shoulders width apart, and (d) sitting with feet shoulders width apart. An example of the distribution for the two first weeks.

the point of feeling the tightness of the hamstring muscles but no pain). Once this position was achieved, the children held it for 20 seconds.

All the participants were urged to maintain their normal levels of physical activity outside of the supervised setting during the research period. Five children in the EG (22%) and seven children in the CG (32%) regularly participated (at least twice per week) in organized extra-curricular sport programmes. During the stretching programme period the EG and CG students participated in traditional games. However, the CG followed the standard PE programme without performing hamstring stretches. Furthermore, the participants in the CG were unaware of the purpose of the study.

Analysis

Descriptive statistics (means and standard deviations) for age, body mass, height, body mass index, and SR scores were calculated. A one-way analysis of variance (ANOVA) was used to study the differences of the general characteristics and pre-intervention SR scores between groups. In addition, chi-squared analyses were carried out to test the ratio differences of gender and extra-curricular sport practitioners between the two groups. Afterward, a two-way ANOVA was applied over the SR scores using the groups (CG, EG) and time factors (pretest, posttest, retest). For the post hoc analyses, α values were corrected using the Bonferroni adjustment. Then, the minimal detectable change at 95% confidence (MDC_{95}) was calculated in order to examine if the change score due to the intervention was true and reliable rather than measurement error (Haley and Fragala-Pinkham, 2006). Moreover, the Hedges' *g* effect size was used to determine the magnitude of treatment effects (Hedges, 2007). Finally, the test–retest reliability of SR test was estimated using both the intraclass correlation coefficient from two-way ANOVA (ICC, including the 95% interval of confidence) (Shrout and Fleiss, 1979) and the Bland and Altman's 95% limits of agreement (LOA) (Bland and Altman, 1986). All statistical analyses were performed using the SPSS version 15.0 for Windows (SPSS® Inc., Chicago, IL). The statistical significance level was set at $p < 0.05$.

Results

All the participants completed the development training programme according to previously established attendance norms. The EG participants obtained an average attendance of 99% in the development training programme. The one-way ANOVA results did not show statistically significant differences in the general characteristics and pretest values between EG and CG ($p > 0.05$). In addition, the chi-square analyses showed that the two groups had a balanced representation of boys and girls ($\chi^2_1 = 0.606$; $p = 0.436$) and a balanced representation of extra-curricular sport practitioners and non-practitioners ($\chi^2_1 = 0.584$; $p = 0.445$).

The results of the two-way ANOVA on the average obtained in the SR showed interaction effects between the group and time variables [$F(2, 86) = 18.625$; $p < 0.001$; $\eta^2_p = 0.302$; $P = 1.000$] (Table 2). For post hoc analysis, the ANOVA with the Bonferroni adjustment showed that the EG increased statistically significant from pretest to posttest ($p < 0.001$). However, for the EG flexibility levels from posttest to retest decreased statistically significant ($p < 0.001$), and statistically differences between the pretest and retest were not found ($p = 1.000$). For the CG no significant differences were found ($p = 1.000$). The MDC_{95} value of the SR score was 1.39 cm, when the average increase in the EG was 2.44 cm. The test–retest reliability for the SR score was very high: ICC, 0.99 (0.98-1.00); LOA, 0.0 $\pm$ 1.7 cm.

Table 2. Effect of the stretching intervention programme on sit-and-reach scores (cm).

Group	Pretest (1) ($M \pm SD$)	Posttest (2) ($M \pm SD$)	Retest (3) ($M \pm SD$)	p[a]	Effect size[b]		
					1–2	2–3	1–3
Experimental ($n = 23$)	16.4 ± 4.9	18.8 ± 5.8***	16.6 ± 5.4***	< 0.001	0.49	−0.41	0.04
Control ($n = 22$)	16.9 ± 5.0	16.9 ± 4.9	16.9 ± 5.2				

Note. M: mean; SD: standard deviation;
[a] Significance level from two-way analysis of variance; post hoc analysis with Bonferroni adjustment; change statistically significant from pretest to posttest, and from posttest to retest (***p < 0.001).
[b] Hedges' g effect size.

Discussion

The first purpose of this study was to evaluate the effects of a 1-minute PE-based stretching programme on the SR score in elementary schoolchildren. The results of the present study show that it is possible to develop children's flexibility by means of an only 1-minute stretching programme in the PE setting. Previous studies in which elementary schoolchildren performed a PE-based stretching programme found a significant improvement in flexibility (Coledam et al., 2012; Rodríguez et al., 1999, 2008; Sainz de Baranda et al., 2006; Santonja et al., 2007).

In the preceding studies the EG students carried out the stretching intervention programme for 5–7 minutes (i.e. from 10–14% of the total session time for a standard class of 50 min) (Coledam et al., 2012; Rodríguez et al., 1999, 2008; Sainz de Baranda et al., 2006; Santonja et al., 2007). Nevertheless, since many curricular contents must be developed in a school year (Ministerio de Educación y Ciencia, 2006), the application of a single curricular content such as the flexibility cannot be allocated a large part of PE time. Therefore, the application of shorter stretching programmes as in the present study could be more suitable in order to facilitate teachers the development of PE contents. The results of the present study showed the effectiveness of a stretching programme carried out for only 1 minute per session (i.e. 2% of the total session time for a standard class of 50 min).

Regarding the magnitude effects of the intervention, the effect size of the present study was moderate indicating that the stretching programme was effective. In contrast with the current results, all the previous studies carried out with elementary schoolchildren obtained large effect sizes ($g = 0.85$–0.94) (Coledam et al., 2012; Rodríguez et al., 1999, 2008; Santonja et al., 2007), except for a study with a very large effects size ($g = 2.06$) (Sainz de Baranda et al., 2006). In addition to the time of stretching per session, other training factors such as the duration of the programme and frequency might clearly influence the magnitude effect of the intervention. Apart from the higher time per session allocated to stretching, the intervention programme of all the above-mentioned studies had a significant longer duration, lasting from 16 weeks (Coledam et al., 2012) to a whole academic year (31–32 weeks) (Rodríguez et al., 1999, 2008; Santonja et al., 2007; Sainz de Baranda et al., 2006). In any case, according to Valentine and Cooper (2003), we have to be aware that in education research even lower values of effect size (0.30) could be considered as of practical relevance. In addition, the results of the MDC_{95} analysis showed that the increase in the SR score due to the intervention programme was true and reliable rather than measurement error.

As regards frequency, although in all the previous studies participants carried out the intervention programme twice a week, Santonja et al. (2007) also found that when the children performed four sessions per week instead of two, the magnitude effect doubled (twice a week, $g = 0.85$; four times a week, $g = 1.53$; both 5 minutes per session, 31 weeks of duration). However, in many countries such as Spain, since PE teachers are limited to a low frequency of two sessions per week (Ministerio de Educación y Ciencia, 2006), the application of intervention programmes with higher frequency is not possible in the PE setting. In this line, the increase of active time for learning in extra-curricular periods would represent an excellent strategy in order for PE teachers to pursue important objectives such as the improvement of flexibility. However, because this strategy mainly depends on the students' autonomy, in many cases such as with the students at these early ages it seems to be impractical.

Another obstacle in PE planning is that after a period of detraining the flexibility gains obtained are expected to decrease (Rancour et al., 2009; Willy et al., 2001). Generally, PE teachers carry out stretching exercises in their PE classes only for a few weeks, and when they cease doing them, they do not know how long the effect will last. Furthermore, a common situation in PE setting is the alternation of academic and holiday periods. Unfortunately, to our knowledge there are no previous studies examining the loss effects of flexibility during detraining periods among schoolchildren. Although most previous studies evaluated the effects of an entire academic year (Rodríguez et al., 1999, 2008; Santonja et al., 2007; Sainz de Baranda et al., 2006) they only examined the global effects, ignoring the changes on flexibility during relatively long detraining periods such as Christmas holiday. Hence, one of the most important outcomes of the present study was the verification that after a detraining period of 5 weeks, children's flexibility reverts back to the baseline level. Therefore, as after 5 weeks of detraining the effect of the stretching programme has completely vanished, PE teachers should continue training students' flexibility after a shorter detraining period in order to maintain the gains obtained in the previous semesters.

For all the above-mentioned reasons, it would be beneficial for PE teachers to know the minimum duration of a flexibility programme to achieve significant improvements, how long it takes to lose the improvements achieved after this development programme, and how a maintenance flexibility programme for schoolchildren should be applied. Thus, a limitation of the present study was a lack of monitoring of the detraining effect in several periods of time (e.g. also after 2 or 3 weeks) or a lack of applying a maintenance programme (Mayorga-Vega et al., 2013b). Future research interventions should examine the effect of different detraining periods among schoolchildren, as well as the application of different maintenance training programmes in order to maintain previous flexibility gains obtained. This knowledge could help the PE teachers to design programmes that guarantee the maintenance of previous flexibility gains. In addition, future studies research should further examine the magnitude effect of different volumes per session or frequency per week.

Finally, in the present study there were some limitations that should be considered when examining its results. The main limitations of this study were related to the validity of the test used. Flexibility is typically characterized by the maximum range of motion in a joint or series of joints (McHugh et al., 1998). Therefore, the angular test that specifically measures hip flexion (straight leg raise test) has been widely considered the criterion measure of hamstring extensibility (Mayorga-Vega et al., 2014b). Nevertheless, as in the present study, when the use of this angular test is limited due to practical issues such as the time constraints, the classic SR test has shown to be a useful alternative to estimate hamstring extensibility (Mayorga-Vega et al., 2014b), even among children at these ages (Hartman and Looney, 2003).

The first factor that could affect the validity of the SR test is the differences in length proportion between the upper and lower limbs (Hoeger et al., 1990). However, since in the present study the pre-intervention values were assessed, any change in the participants' flexibility was compared with their baseline level. In addition, another potential limitation of using a lineal test might be due to the change in the proportion of the anthropometric parameters and/ or in the position of the spine between the pretest and the following measures in the posttest and retest. Nevertheless, besides the relatively short period between the measures taken, the study design included an equivalent CG (i.e. characteristics such as age, height, body mass, gender ratio, extra-curricular sport practitioners ratio or pretest values were equivalent between groups). Therefore, we could assume reasonably that any change in the proportion of the anthropometric parameters and/ or in the position of the spine should be similar in both groups.

In conclusion, to our knowledge this study is the first one that examines the effect of a 1-minute stretching programme and its subsequent detraining on SR score among schoolchildren. The results of the present study suggest that it is possible to develop schoolchildren's flexibility through an only 1-minute stretching programme in the PE setting. However, after a 5-week detraining period, children revert back to their initial flexibility levels. Hence, due to the fact that after 5 weeks of detraining the effect of the stretching programme has completely vanished, PE teachers should continue exercising students' flexibility after a shorter detraining period in order to maintain the gains obtained in the previous semesters.

Acknowledgements

We thank Anna Szczesniak for the English revision. Results were prior published, in abstract form, in the proceedings of 6th Baltic Scientific Conference Sport Science for Sustainable Society April 23-25, 2013 Riga, Latvia.

Funding

This work was supported by the Spanish Ministry of Education, Culture and Sport [grant number AP2010-5905 (second author)].

References

Bland JM and Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327(8476): 307–310.

Carregaro RL and Coury HJC (2009) Does reduced hamstring flexibility affect trunk and pelvic movement strategies during manual handling? *International Journal of Industrial Ergonomics* 39(1): 115–120.

Chiodera P, Volta E, Gobbi G, et al. (2008) Specifically designed physical exercise programmes improve children's motor abilities. *Scandinavian Journal of Medicine & Science in Sports* 18(2): 179–187.

Coledam DHC, Arruda GA and Ramos de Oliveira A (2012) Chronic effect of static stretching performed during warm-up on flexibility in children. *Revista Brasileira de Cineantropometria e Desempenho Humano* 14(3): 296–304.

Feldman DE, Shrier I, Rossignol M, et al. (2001) Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology* 154(1): 30–36.

Haley SM and Fragala-Pinkham MA (2006) Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy* 86(5): 735–743.

Hartman JG and Looney M (2003) Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science* 7(2): 71–87.

Hedges LV (2007) Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics* 32(4): 341–370.

Hoeger WW, Hopkins DR, Button S, et al. (1990) Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science* 2(2): 156–162.

Jones MA, Stratton G, Reilly T, et al. (2005) Biological risk indicators for recurrent non-specific low back pain in adolescents. *British Journal of Sports Medicine* 39(3): 137–140.
Kanásová J (2008) Reducing shortened muscles in 10–12-year-old boys through a physical exercise programme. *Medicina Sportiva* 12(4): 115–123.
Kang MH, Jung DH, An DH, et al. (2013) Acute effects of hamstring-stretching exercises on the kinematics of the lumbar spine and hip during stoop lifting. *Journal of Back and Musculoskeletal Rehabilitation* 26(3): 329–336.
Kendall FP, McCreary EK, Provance PG, et al. (2005) *Muscles: Testing and function with posture and pain*. 5th edition. Baltimore: Lippincott, Williams, & Wilkins.
López-Miñarro PA, Muyor JM and Alacid F (2012) Influence of hamstring extensibility on sagittal spinal curvatures and pelvic tilt in high-trained young kayakers. *European Journal of Sport Science* 12(6): 469–474.
López-Miñarro PA, Sainz de Baranda P and Rodríguez-Garcia PL (2009) A comparison of the sit-and-reach test and the back-saver sit-and-reach test in university students. *Journal of Sports Science & Medicine* 8(1): 116–122.
McHugh MP, Kremenic IJ, Fox MB, et al. (1998) The role of mechanical and neural restraints to joint range of motion during passive stretch. *Medicine and Science in Sports and Exercise* 30(6): 928–932.
Mayorga-Vega D, Merino-Marban R and Garcia-Romero JC (2013a) Validity of sit-and-reach with plantar flexion test in children aged 10-12 years. *Revista Internacional de Medicina y Ciencias de la Actividad Fısica y del Deporte*. Epub ahead of print 21 March 2013.
Mayorga-Vega D, Merino-Marban R, Garrido FJ, et al. (2014a) Comparison between warm-up and cool-down stretching programmes on hamstring extensibility gains in primary schoolchildren. *Physical Activity Review* 2: 16–24.
Mayorga-Vega D, Merino-Marban R and Viciana J (2014b) Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine* 13(1): 1–14.
Mayorga-Vega D, Viciana J and Cocca A (2013b) Effects of a circuit training programme on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics* 37: 153–160.
Mikkelsson L, Nupponen H, Kaprio J, et al. (2006) Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine* 40(2): 107–113.
Ministerio de Educación y Ciencia (2006). Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria. [Royal Decree 1513/2006, the 7th December, laying down the Curriculum for Primary Education.] *Government Gazette*, 293, 43075–43080.
Muyor JM, Alacid F, Rodríguez-García PL, et al. (2012) Influence of hamstring extensibility on sagittal spinal curvatures and pelvic inclination in athletes. *International Journal of Morphology* 30(1): 176–181.
National Association for Sports and Physical Education (2005) *Physical Education for Lifelong Fitness: The Physical Best Teacher's Guide*. Leeds: Human Kinetics.
Peiró Velert C and Devis Devís J (1995) Health-based physical education in Spain: The conception, implementation and evaluation of an innovation. *European Physical Education Review* 1(1): 37–54.
Ramos Espada D, González Montesinos JL and Mora Vicente J (2007a) Differences in joint amplitudes among males and females of school age. *Apunts Medicina de l'Esport* 42(1): 13–25.
Ramos Espada D, González Montesinos JL and Mora Vicente J (2007b) Evolution of the articular range in primary and secondary school. *Revista Internacional de Medicina y Ciencias de la Actividad Fısica y del Deporte* 7(26): 144–157.
Rancour J, Holmes CF and Cipriani DJ (2009) The effects of intermittent stretching following a 4-week static stretching protocol: A randomized trial. *Journal of Strength & Conditioning Research* 23(8): 2217–2222.
Rodríguez PL, Santonja F, Canteras M, et al. (1999) Improvement of hamstring extensibility after school programme of stretching. *Selección* 8(4): 157–164.

Rodríguez PL, Santonja FM, López-Miñarro PA, et al. (2008) Effect of physical education stretching programme on sit-and-reach score in schoolchildren. *Science & Sports* 23(3-4): 170–175.

Sainz de Baranda P, Rodríguez PL, Santonja FM, et al. (2006) Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *Journal of Human Movement Studies* 51(4): 277–289.

Santonja F, Rodríguez PL, Sainz de Baranda P, et al. (2004) Role of physical education teacher to the misalignment of the spine. *Selección* 13(1): 5–17.

Santonja FM, Sainz De Baranda P, Rodríguez PL, et al. (2007) Effects of frequency of static stretching on straight-leg raise in elementary school children. *Journal of Sports Medicine and Physical Fitness* 47(3): 304–308.

Shrout PE and Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2): 420–428.

Siedentop D and Tanehill D (2000) *Developing Teaching Skills in Physical Education*. Mountain View: Mayfield Publishing Company.

Sjölie AN (2004) Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scandinavian Journal of Medicine & Science in Sports* 14(3): 168–175.

Thacker SB, Gilchrist J, Stroup DF, et al. (2004) The impact of stretching on sports injury risk: A systematic review of the literature. *Medicine and Science in Sports and Exercise* 36(3): 371–378.

Valentine J and Cooper H (2003) *Effect Size Substantive Interpretation Guidelines: Issues in the Interpretation of Effect Sizes*. Washington: What Works Clearing House.

Willy RW, Kyle BA, Moore SA, et al. (2001) Effect of cessation and resumption of static hamstring muscle stretching on joint range of motion. *Journal of Orthopaedic and Sports Physical Therapy* 31(3): 138–144.

Author biographies

Rafael Merino-Marban is a Professor in the Faculty of Educational Sciences at the University of Málaga, Spain.

Daniel Mayorga-Vega is a Research fellow in the Department of Physical Education and Sport at the University of Granada, Spain.

Emilio Fernandez-Rodríguez is a Professor in the Faculty of Educational Sciences at the University of Málaga, Spain.

Francisco Vera Estrada is a Physical Education Teacher in the Elementary Education School Tartessos, Málaga, Spain.

Jesús Viciana is a Professor in the Faculty of Physical Education and Sport in the University of Granada, Spain.

Paper XIV

EFFECTS OF A CIRCUIT TRAINING PROGRAM ON MUSCULAR AND CARDIOVASCULAR ENDURANCE AND THEIR MAINTENANCE IN SCHOOLCHILDREN

Mayorga-Vega, D., Viciana, J., & Cocca, A.

Journal of Human Kinetics

2013, *37*, 153-160.

Journal of Human Kinetics volume 37/2013, 153-160 DOI:10.2478/hukin-2013-0036 153
Section III – Sports Training

Effects of a Circuit Training Program on Muscular and Cardiovascular Endurance and their Maintenance in Schoolchildren

by
Daniel Mayorga-Vega[1], Jesús Viciana[2], Armando Cocca[3]

The purpose of this study was to evaluate the effects of a circuit training program along with a maintenance program on muscular and cardiovascular endurance in children in a physical education setting. Seventy two children 10-12 years old from four different classes were randomly grouped into either an experimental group (n = 35) or a control group (n = 37) (two classes for each group). After an eight-week development program carried out twice a week and a four-week detraining period, the experimental group performed a four-week maintenance program once a week. The program included one circuit of eight stations of 15/45 to 35/25 seconds of work/rest performed twice. Abdominal muscular endurance (sit-ups in 30 seconds test), upper-limbs muscular endurance (bent arm hang test), and cardiovascular endurance (20-m endurance shuttle run test) were measured at the beginning and at the end of the development program, and at the end of the maintenance program. After the development program, muscular and cardiovascular endurance increased significantly in the experimental group (p < 0.05). The gains obtained remained after the maintenance program. The respective values did not change in the control group (p > 0.05). The results showed that the circuit training program was effective to increase and maintain both muscular and cardiovascular endurance among schoolchildren. This could help physical education teachers design programs that permit students to maintain fit muscular and cardiovascular endurance levels.

Key words: *physical fitness program, health-related physical fitness, muscular strength, cardiorespiratory fitness, school-based program, physical education.*

Introduction

Physical fitness is nowadays considered as one of the most important health markers in childhood (Ortega et al., 2008). Consequently, in the last decades several countries have been promoting physical fitness improvement among young people in different ways (Department of Health and Human Services, 1990). In many circumstances, schools have been considered the best setting in which children with low fitness levels can be identified and a healthy lifestyle can be promoted (Ortega et al., 2008). Therefore, one of the main Spanish government strategies was focused on modifying school legislations in order to give health a more important role in the Educational System (Ministerio de Educación y Ciencia, 2006). Schools are mainly attempting to increase the pupils' health level by using measures such as the improvement of their physical fitness through physical education (PE) (Ministerio de Educación y Ciencia, 2006). It has been concluded that the health promotion policies and physical activity programs should be

[1] - *Department of Physical Education and Sport, University of Granada, Spain.*
[2] - *Department of Physical Education and Sport, University of Granada, Spain.*
[3] - *Department of Physical Education and Sport, University of Granada, Spain.*

Authors submitted their contribution of the article to the editorial board.
Accepted for printing in Journal of Human Kinetics vol. 37/2013 on June 2013.

designed to improve physical fitness, where strength and cardiovascular endurance are the most important health-related physical fitness components (Ortega et al., 2008).

It is known that planning long-term fitness programs is the best way to improve these components (Donnelly et al., 2009). Nonetheless, in the PE setting these programs cannot last the whole course or a large part of it since many curricular contents must be developed in a school year (Ministerio de Educación y Ciencia, 2006). Consequently, in the PE setting we need to find short-term programs that could be also effective for the increment of fitness. One of the methodologies that meet these criteria could be the circuit training (Dorgo et al., 2009; Granacher et al., 2011a; Granacher et al., 2011b). The circuit training effectively reduces the time devoted to training while allowing an adequate training volume to be achieved (Alcaraz Ramón et al., 2008). Moreover, it permits a greater motor engagement time (Lozano et al., 2009), which is a very important requirement for the success of a PE program. In addition, this methodology has multilevel effects on fitness, especially in beginners (Alcaraz Ramón et al., 2008; Dorgo et al., 2009; Wong et al., 2008).

Other problem related to physical fitness is its expected decrease after a period of detraining. Several authors confirm that after 8 to 12 weeks of detraining children lose a significant part of the physical fitness gains obtained (Da Fontoura et al., 2004; Faigenbaum et al., 1996; Ingle et al., 2006; Isaacs et al., 1994; Tsolakis et al., 2004). A possible solution for this problem could be the periodical introduction of short maintaining programs throughout the academic course. During these intervals, PE teachers would be able to develop other curricular contents and at the same time they could be improving the previous physical fitness gains. These programs could permit to keep the physical fitness level achieved without interfering in the normal course of the PE planning. Unfortunately, to our knowledge there are no studies addressing the effect of a physical fitness maintenance program in a PE setting. Consequently, the purpose of this study was to evaluate the effects of a circuit training program and its maintenance on muscular and cardiovascular endurance levels in children in a PE setting.

Material and Methods

Participants

Seventy two apparently healthy Spanish children (boys, n = 40; girls, n = 32) 10-12 years old (age 11.10 ± 0.38 years; body mass 43.29 ± 10.45 kg; body height 1.43 ± 0.07 m; body mass index 19.39 ± 3.90 kg/m^2) from four different PE classes of a public primary school participated in this study. For practical reasons and the nature of the present study (the intervention was focused on natural groups in a school context) a cluster randomized controlled trial was used. General profiles were assigned randomly to form one of the study groups (two classes for each group): experimental group (EG, n = 35), or control group (CG, n = 37). EG and CG consisted of two gender balanced groups (47% and 43% of girls, respectively).

All participants were urged to maintain their normal levels of physical activity outside of the supervised setting. Twenty-seven children in the EG (77%) and 27 children in the CG (73%) regularly participated (at least twice per week) in organized sports programs. Children and their legal guardians were fully informed about all the features of the study and were required to sign an informed consent form. The Ethical Committee of the University of Granada approved the study protocol.

Measures

The participants were evaluated using the muscular and cardiovascular endurance tests included in the EUROFIT battery (Council of Europe Committee for the Development of Sport, 1988), validated and standardized by the Council of Europe. The test sessions were carried out during the PE classes at the beginning and at the end of the development circuit training program (pretest and posttest), in order to see the changes that were produced. Subsequently, after a period of detraining and the application of the maintenance circuit training program, the participants were evaluated again (retest).

The tests were administered in an indoor sports center court with a non-slip floor, under the same environmental conditions, on the same day and at the same time for each student. A blind evaluation was carried out by two researchers following the standard protocol for each test. Each researcher assessed physical fitness with the same tests using identical equipment. Prior to the evaluation, the participants completed a standardized warm-up consisting of five minutes of running from low to moderate intensity. The order and a brief description of the test protocol are as follows:

Sit-ups in 30 seconds test (SUP). This test was used to measure abdominal muscular endurance. The participants laid supine on the mat with their knees flexed at an angle of 90 degrees and their feet flat on the floor, stabilized by a researcher. The participant´s fingers were to be interlocked behind their head. On the command 'Go', the participants´ elbows had to contact the knees and return to the starting position as many times as possible in 30 s. Each participant was allowed to perform the test once. The total number of sit-ups performed in 30 s was recorded.

Bent arm hang test (BAH). This test was used to measure upper-limbs muscular endurance. The participants had to maintain a bent arm position while hanging from a bar with hands in a pronated grip and at shoulder width. The participants' chin had to be above the bar and held in this position as long as possible. The test ended when the participants' eyes went below the bar. Each participant was allowed to perform the test once. The total time in seconds was retained.

20-m endurance shuttle run test (ESR). This test was used to assess cardiovascular endurance. All students ran between two parallel lines put 20 m apart, in the rhythm marked by a recorded beep. A researcher ran alongside the children to help them keep the desired pace. The starting speed was 8.5 km/h; and it increased 0.5 km/h every minute. The test ended when the child stopped running due to fatigue or failed to reach the line before the next signal for two consecutive times. Each participant was allowed to perform the test once. The last completed lap (timed in seconds) was recorded.

Procedures

A circuit training program was applied to the EG during the PE classes under the supervision of a researcher. Firstly, the EG participants performed a development circuit training program twice a week on nonconsecutive days for eight weeks. They completed a total of 14 training sessions, since two classes coincided with festivals and could not be used. Then, after a period of detraining (four weeks) coinciding with Christmas holiday, the EG participants completed a maintenance circuit training program one session per week during four weeks. During the period of maintenance program, each session of maintenance was alternated with a normal class of PE according to the course planning designed by the teacher.

Each session lasted 50 minutes and consisted of a five-minute warm up during which children had to play a racing game, 40-minute circuit training, and two series of a 15-30 second cool-down of static stretching, primarily for the hamstrings and lumbar region (Table 1). All exercises were fully explained and previously demonstrated by the researcher, and children were asked to try them during a few minutes before starting the first session of the intervention. According to previous studies carried out in the PE setting (Dorgo et al., 2009; Granacher et al., 2011a; Granacher et al., 2011b), the intervention was organized in a circuit program. One circuit of eight stations was developed, and then repeated twice in each session. Each station consisted of an exercise lasting from 15 to 35 seconds (extended progressively from the first session to the last), and the rest time between them was of 45-25 seconds (gradually reduced during the program). The increase of the work time and the decrease of the rest time along the intervention were based on the training load progression principle.

During the work time the students should complete as many repetitions as possible in a controlled manner. As other studies show, the last repetition of each set represents the momentary muscular fatigue (Faigenbaum et al., 2002; Faigenbaum et al., 2005). In order to achieve it, the children were offered three levels of difficulty in each station (Table 1), so that the intensity of exercise was best suited to each student. All participants began at the first level of difficulty, and when a student could perform more than one repetition per second, he/she was allowed to advance to the next level. With the aim of developing cardiovascular endurance, at the end of each circuit all students simultaneously executed an additional stage consisting of a five-minute endurance racing game. The researcher gave positive feedback to motivate participants in achieving it (Badami et al., 2011).

Table 1

Circuit training session

Phase (time)/ Exercises	Intensive progression (level 1/ 2/ 3)[a]	Material
Warm-up (5 min)		
Racing games		
Main part (40 min)		
Circuit training stations		
a. Throwing from chest	1kg/ 1.5kg/ 2kg	MB
b. Rowing	Low/ medium/ high resistance	Elastic band
c. Going up-down	Body weight/ +1kg/ +2kg	Swedish bench, MB
d. Triceps extension	Low/ medium/ high resistance	Elastic band
e. Biceps curl	Low/ medium/ high resistance	Elastic band
f. Skipping rope	Micropause/ with/ without rebound	Rope
g. Crunches	Arms stretched forward/ chest/ backward	Mat
h. Bridging	Body weight/ +1kg/ +2kg	Mat, MB
Additional station		
i. Racing games		
Cold-down (5 min)		
Static stretching		

MB = Medicine ball; [a] All participants began at the first level of difficulty. When a student could perform more than one repetition per second was allowed to advance to next level.

During the development and maintenance programs the EG executed the circuit training,while the CG participated in traditional games, basketball and volleyball activities. However, during the maintenance program the EG alternated one session of physical fitness maintenance with other activities such as basketball and volleyball. No participant was allowed to carry out any physical fitness training outside of the supervised setting.

Analysis

Descriptive statistics (means and standard deviations) for age, body height, body mass index, and muscular and cardiovascular endurance results were calculated. The Student´s t test for independent samples was used to study the differences of the general characteristics between groups. As the BAH variable did not follow a normal distribution, the data was transformed using a logarithm (Bland and Altman, 1996). Because a higher precision was required for ESR test performance, the final time spent in the test was expressed in seconds, instead of stages or half stages, and it was used for statistical analysis (Ruiz et al., 2011). A two-way analysis of variance (ANOVA) was applied over the dependent variables (SUP, BAH, ESR) using groups (EG, CG) and time factors (pretest, posttest, retest). For the post-hoc analyses, α values were corrected using the Bonferroni adjustment. The Hedges´ g effect size was used to determine the magnitude of treatment effects (Hedges, 2007). The test-retest reliability for muscular and cardiovascular endurance tests was estimated using the intraclass correlation coefficient from two-way ANOVA (ICC3,k) (Shrout and Fleiss, 1979). Furthermore, 95% interval of confidence was calculated. All statistical analyses were performed using the SPSS version 15.0 for Windows (SPSS® Inc., Chicago, IL). The statistical significance level was set at $p < 0.05$.

Results

All students completed the development training program and 67 the maintenance training program according to previously established norms (no more than two classes were missed in the development training program, and none were missed in the maintenance training). Retest data of four participants from the EG and one from the CG were excluded due to missed classes in the maintenance training program and absence in the retest session test, respectively. The EG participants finally considered for analysis obtained an average attendance of 94% and 100% in the development and maintenance training program, respectively. The Student´s t for independent samples results did not show statistically significant differences in the general characteristics between EG and CG.

Sit-ups in 30 seconds test. The EG had significantly greater gains in SUP compared to the CG [F(2, 63) = 4.636; p = 0.011; η^2_P = 0.069; P = 0.773] (Table 2). The ANOVA with Bonferroni adjustment showed that the EG increased significantly from pretest to posttest (p = 0.026) and from pretest to retest (p = 0.004). Nevertheless, the difference from the posttest to the retest for the EG was not statistically significant (p = 0.105). No significant differences were found for the CG (p = 1.000). The test-retest reliability for the SUP was 0.86 (0.73-0.93).

Bent arm hang test. Significantly greater gains were found for the EG compared to CG [F(2, 63) = 5.994; p = 0.003; η^2_P = 0.087; P = 0.875]. The EG participants significantly increased FAH from pretest to posttest (p = 0.009) and from pretest to retest (p < 0.001). For the EG, the improvement from the posttest to the retest approached statistical significance (p = 0.065). No differences were found for the CG (p ≥ 0.324). The test-retest reliability for the BAH was 0.95 (0.90-0.97).

20-m endurance shuttle run test. The EG had significantly greater gains in ESR compared to the CG [F(2, 64) = 5.230; p = 0.007; η^2_P = 0.076; P = 0.824]. The ANOVA with Bonferroni adjustment showed that the EG increased significantly from pretest to posttest (p = 0.015) though improvement from the pretest to the retest approached statistical significance (p = 0.088). No significant differences were found from posttest to retest for the EG (p = 0.210). No differences were found for CG (p ≥ 0.975). The test-retest reliability for the ESR was 0.90 (0.81-0.95).

Discussion

The results of the present study show that it is possible to develop both muscular and cardiovascular endurance by means of an eight-week circuit training program in the PE setting. Previous studies in which children performed an extra-curricular circuit training program confirmed a significant improvement on both muscular and cardiorespiratory fitness (Annesi et al., 2005; Ignico and Mahon, 1995; Wong et al., 2008).

Previous studies in which children performed an extra-curricular circuit training program confirmed a significant improvement on both muscular and cardiorespiratory fitness (Annesi et al., 2005; Ignico and Mahon, 1995; Wong et al., 2008). Nevertheless, the design and the procedure of the present study depended on many aspects related to the school context as previously discussed in this manuscript. Likewise, due to the lack of special machines in a PE setting, in the present study body weight, elastic band and ball exercises (Annesi et al., 2005; Faigenbaum and Mediate, 2008; Flanagan et al., 2002) were used instead of specific strength equipment (Granacher et al., 2011a; Granacher et al., 2011b).

One of the main objectives of the PE teachers at these educational levels is to make the pupils active as long as possible during the classes. With the circuits method the pupils can easily reach the minimum motor engagement time (Lozano et al., 2009) at the same time they execute many types of exercises. This is the best way to make the most of the time at a PE teacher's disposal, especially when classes are few and short-lasting and there are many contents to develop (Ministerio de Educación y Ciencia, 2006). Thus, the present results indicate that the design proposed in this research could be effective for PE classes. In this line, Dorgo et al. (2009) carried out a circuit training program with adolescents in the PE setting. These authors found a statistically significant improvement for both muscular strength and cardiovascular endurance when the circuit training was complemented with endurance training.

One of the most important outcomes of this study was that a maintenance program carried out once a week in four weeks could be effective to maintain the gains previously obtained. As explained before, the majority of studies coincide in eight weeks setting as the period of inactivity determining the complete loss of previous physical fitness gains (Faigenbaum et al., 1996; Isaacs et al., 1994; Tsolakis et al., 2004).

Table 2

Muscular and cardiovascular endurance performance for the development and maintenance circuit training program

Group	Pretest (1) (M ± SD)	Posttest (2) (M ± SD)	Retest (2) (M ± SD)	p	Effect size		
					1-2	2-3	1-3
SUP (nº)							
Experimental	20.37 ± 4.21	22.09 ± 3.70*	23.10 ± 3.88††	<0.05	0.44	0.22	0.68
Control group	17.95 ± 4.94	17.64 ± 5.89	17.57 ± 5.76				
BAH[a] (s)							
Experimental	11.63 ± 9.93	14.65 ± 12.75**	20.80 ± 19.07†††	<0.01	0.38	0.31	0.65
Control group	16.87 ± 18.84	14.13 ± 14.09	16.11 ± 15.95				
ESR (s)							
Experimental	160.71 ± 94.33	198.71 ± 98.43*	186.00 ± 84.76	<0.01	0.48	-0.11	0.37
Control group	191.78 ± 96.66	184.06 ± 96.12	181.78 ± 100.22				

M = Mean; SD = Standard deviation; SUP = Sit-ups in 30 seconds test; BAH = Bent arm hang test; ESR = 20-m endurance shuttle run test; BAH[a] = for statistical analysis the raw data were transformed by the logarithm; p = significance level from two-way analysis of variance; Effect size = Hedges´g effect size.
Post-hoc analyses with Bonferroni adjustment:
*Change statistically significant from pretest to posttest (*p < 0.05, **p < 0.01);*
Change statistically significant from pretest to retest (††p < 0.01, †††p < 0.001)

In the present study the sum of the periods of detraining and maintenance was eight weeks, thus an unsatisfactory design of the maintenance program should have matched a decrease (or the complete dissipation) of the physical fitness profits. Nevertheless, results were positive since the muscular and cardiovascular endurance was maintained after these weeks.

In line with the present study, DeRenne et al. (1996) found out that a maintenance program carried out once a week in pubescent basketball players was efficient to retain strength. However, Blimkie et al. (1989) found that a maintenance program carried out once a week in pre-pubescent children was not efficient to retain strength. Unfortunately, previous studies that examined the maintenance of cardiovascular endurance in youth were not found. In addition, the previous studies were carried out in an extra-curricular period and not in a PE setting. Furthermore, the researchers applied the maintenance program just after the training program, that is, without a period of inactivity between development and maintenance. In the present study, a maintenance program was applied after a period of detraining because it is the most common situation in normal PE planning (due to the typical alternation of holidays, academic periods and the need to teach other curricular contents in the PE classes). Consequently, the design of the present study seems to be suitable for the school environment as it respects all the features and norms established in it. Moreover, it should be effective for increasing the strength and cardiovascular endurance values and then maintaining them during larger periods.

In conclusion, the present study suggests that it is possible to develop and maintain muscular and cardiovascular endurance through a short-term program in the PE setting. Maintenance programs appear to be necessary in the school context to make the physical fitness training effective and feasible within a school plan, permitting at the same time the regular development of other curricular contents. Their utilization could permit the PE teacher to design programs that guarantee the maintenance of previous muscular and cardiovascular endurance gains in a few sessions. Even though more research is needed to confirm these results, the maintenance program could become a principal element to normal PE planning in the future. A limitation of the present study was the fact that the intervention program did not include many playful tasks. At these ages in PE classes it is important to develop contents mainly through ludic activities. Future interventions should focus on physical fitness programs based on stations with games, as well as the effect of the combination of different frequencies and durations of maintenance training.

Acknowledgements

We thank Anna Szczesniak for the English revision. Daniel Mayorga-Vega is supported by a research grant from the Spanish Ministry of Education (AP2010-5905).

References

Alcaraz Ramón PE, Sánchez-Lorente J, Blazevich AJ. Physical performance and cardiovascular responses to an acute bout of heavy resistance circuit training versus traditional strength training. *J Strength Cond Res*, 2008; 22: 667-671

Annesi J, Westcott W, Faigenbaum A, Unruh J. Effects of a 12 week physical activity protocol delivered by YMCA after-school counsellors (Youth Fit for Life) on fitness and self-efficacy changes in 5-12 year old boys and girls. *Research Q Exercise Sport*, 2005; 76: 468-476

Badami R, VaezMousavi M, Wulf G, Namazizadeh M. Feedback after good versus poor trials affects intrinsic motivation. *Research Q Exercise Sport*, 2011; 82: 360-364

Bland JM, Altman D. Transforming data. *Brit Med J*, 1996; 312: 770

Blimkie CJ, Martin J, Ramsay J, Sale D, MacDougall D. The effects of detraining and maintenance weight training on strength development in prepubertal boys. *Can J Appl Sport Sci*, 1989; 14: S102

Council of Europe Committee for the Development of Sport. *Eurofit: Handbook for the EUROFIT tests of physical fitness*. Italy: Edigraf editorial grafica; 1988

Da Fontoura AS, Schneider P, Meyer F. Effect of the muscular strength detraining in prepubertal boys. *Rev Bras Med Esporte*, 2004; 10: 285-288

Department of Health and Human Services. *Healthy people 2000: National health promotion and disease prevention objectives*. Washington, DC: DHHS Publ No (PHS), 91-50212; 1990

DeRenne C, Hetzler RK, Buxton BP, Ho KW. Effects of training frequency on strength maintenance in pubescent baseball players. *J Strength Cond Res*, 1996; 10: 8-14

Donnelly J, Greene J, Gibson C, Smith B, Washburn R, Sullivan D, DuBose K, Mayo MS, Schmelzle KH, Ryan JJ, Jacobsen DJ, Williams SL. Physical activity across the curriculum (PAAC): A randomized controlled trial to promote physical activity and diminish overweight and obesity in elementary school children. *Prev Med*, 2009; 49: 336-341

Dorgo S, King GA, Candelaria NG, Bader JO, Brickey GD, Adams CE. Effects of manual resistance training on fitness in adolescents. *J Strength Cond Res*, 2009; 23: 2287-2294

Faigenbaum AD, Mediate P. Medicine ball training for kids. *Health Fit J*, 2008; 12: 7-12

Faigenbaum AD, Milliken LA, LaRosa-Loud R, Burak BT, Doherty CL, Westcott WL. Comparison of 1 and 2 days per week of strength training in children. *Research Q Exercise Sport*, 2002; 73: 416-424

Faigenbaum AD, Milliken LA, Moulton L, Westcott WL. Early muscular fitness adaptations in children in response to two different resistance training regimens. *Pediatr Exerc Sci*, 2005; 17: 237-248

Faigenbaum AD, Westcott WL, Micheli LJ, Outerbridge AR, Long CJ, LaRosa-Loud R, Zaichkowsky LD. The effects of strength training and detraining on children. *J Strength Cond Res*, 1996; 10: 109-114

Flanagan SP, Laubach LL, De Marco GM, Alvarez C, Borchers S, Dressman E, Gorka C, Lauer M, Mckelvy A, Metzler M, Poeppelman J, Redmond C, Riggenbach M, Tichar S, Wallis K, Weseli D. Effects of two different strength training modes on motor performance in children. *Research Q Exercise Sport*, 2002; 73: 340-344

Granacher U, Goesele A, Roggo K, Wischer T, Fischer S, Zuerny C, Gollhofer A, Kriemler S. Effects and mechanisms of strength training in children. *Int J Sports Med*, 2011a; 32: 357-364

Granacher U, Muehlbauer T, Doerflinger B, Strohmeier R, Gollhofer A. Promoting strength and balance in adolescents during physical education: Effects of a short-term resistance training. *J Strength Cond Res*, 2011b; 25: 940-949

Hedges LV. Effect sizes in cluster-randomized designs. *J Educ Behav Stat*, 2007; 32: 341-370

Ignico AA, Mahon AD. The effects of a physical fitness program on low-fit children. *Research Q Exercise Sport*, 1995; 66: 85-90

Ingle L, Sleap M, Tolfrey K. The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *J Sport Sci*, 2006; 24: 987-997

Isaacs L, Pohlman R, Craig B. Effects of resistance training on strength development in prepubescent females. *Med Sci Sport Exer*, 1994; 26: S210

Lozano L, Viciana J, Martínez JC, Cocca A, Jiménez R. Influence of classroom environment and motor engagement time. *Rev Mex Psicol*, 2009; 26: 675-676

Ministerio de Educación y Ciencia. Royal decree 1513/2006, the 7th December, laying down the Curriculum for Primary Education. Madrid, Spain: Government Gazette, 43053-43102; 2006

Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: A powerful marker of health. *Int J Obesity*, 2008; 32: 1-11

Ruiz JR, España Romero V, Castro Piñero J, Artero EG, Ortega FB, Cuenca García M, Jiménez Pavón D, Chillón P, Girela Rejón MJ, Mora J, Gutiérrez A, Suni J, Sjöstrom M, Castillo MJ. APHA-Fitness battery: Field test for assessing health-related fitness among children and adolescents. Appendix. *Nutr Hosp*, 2011; 26: 1210-1214

Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*, 1979; 86; 420-428

Tsolakis CK, Vagenas GK, Dessypris AG. Strength adaptations and hormonal responses to resistance training and detraining in preadolescent males. *J Strength Cond Res*, 2004; 18: 625-629

Wong PCH, Chia MYH, Tsou IYY, Wansaicheong GKL, Tan B, Wang JCK, Tan J, Kim C, Boh G, Lim D. Effects of a 12-week exercise training programme on aerobic fitness, body composition, blood lipids and C-reactive protein in adolescents with obesity. *Ann Acad Med Singapore*, 2008; 37: 286-293

Corresponding author:
Daniel Mayorga-Vega
Department of Physical Education and Sport, University of Granada
Alfacar street, 18011, Granada (Spain)
Phone: 0034-958-24-6641
Fax r: 0034-958-24-43-69
E-mail: dmayorgavega@gmail.com

Paper XV

EFFECTS OF A PHYSICAL EDUCATION-BASED PROGRAM ON HEALTH-RELATED PHYSICAL FITNESS AND ITS MAINTENANCE IN HIGH SCHOOL STUDENTS: A CLUSTER-RANDOMIZED CONTROLLED TRIAL

Mayorga-Vega, D., Montoro-Escaño, J., Merino-Marban, R., & Viciana, J.

European Physical Education Review

In press

Original Article

EPER

European Physical Education Review
1–17

DOI: 10.1177/1356336X15599010
epe.sagepub.com

SAGE

Effects of a physical education-based programme on health-related physical fitness and its maintenance in high school students: A cluster-randomized controlled trial

Daniel Mayorga-Vega
Department of Physical Education and Sport, University of Granada, Granada, Spain

Jorge Montoro-Escaño
IES Torre Atalaya, Junta de Andalucía, Málaga, Spain

Rafael Merino-Marban
Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Málaga, Spain

Jesús Viciana
Department of Physical Education and Sport, University of Granada, Granada, Spain

Abstract

The purpose of this study was to examine the effects of a physical education-based development and maintenance programme on objective and perceived health-related physical fitness in high school students. A sample of 111 students aged 12–14 years old from six classes were cluster-randomly assigned to an experimental group ($n = 54$) or a control group ($n = 57$). During the physical education classes, the experimental group students performed a development programme twice a week for nine weeks. Then, after a four-week period of detraining, the experimental group students completed a maintenance programme twice a week for eight weeks. Students' objective cardiorespiratory fitness (20-meter shuttle run test), objective muscular fitness (flexion-extension

Corresponding author:
Daniel Mayorga-Vega, Department of Physical Education and Sport, University of Granada, Alfacar street, 18011, Granada, Spain.
Email: dmayorgavega@gmail.com

legs test), and perceived physical fitness (International Fitness Scale and the Contour Drawing Rating Scale) were measured at the beginning and at the end of the development programme, as well as at the end of the maintenance programme. The results of the two-way analyses of variance showed that the physical education-based development programme significantly increased the students' objective physical fitness levels, and that these gains obtained remained after the maintenance programme ($p < 0.01$). However, the intervention programme did not show an influence on the students' perceived physical fitness ($p > 0.05$). The physical education-based intervention programme was effective in increasing and maintaining objective physical fitness among high school students. This knowledge could help teachers to design programmes that permit a feasible and effective development and maintenance of health-related physical fitness in a physical education setting.

Keywords

Intervention programme, cardiorespiratory fitness, muscular fitness, perceived physical fitness, adolescents

Received January 29, 2015. Accepted for publication July 14, 2015.

Introduction

Physical fitness is nowadays considered as one of the most powerful markers of health, even above other traditional markers such as weight status, blood pressure or cholesterol level (Blair, 2009). Current evidence has shown how physical fitness status is an important predictor of morbidity and mortality in adults (Kodama et al., 2009). Particularly in children and adolescents, for instance, higher cardiorespiratory fitness levels have been associated with a healthier cardiovascular profile in adulthood, and muscular strength improvements have been negatively related to changes in overall adiposity (Ruiz et al., 2009). Furthermore, in young people higher physical fitness levels have been associated with a healthier psychological status related to markers such as depression, anxiety, self-esteem or self-concept (Ortega et al., 2008; Ruiz et al., 2009).

Unfortunately, currently low physical fitness affects a large number of adolescents. For instance, in Spain about one in five adolescents has a physical fitness level indicative of future cardiovascular risk (Ortega et al., 2005). Therefore, health promotion policies should be designed to promote physical fitness levels from childhood, especially cardiorespiratory and muscular fitness, which are the most important health-related physical fitness components (Ortega et al., 2008; Ruiz et al., 2009). Regarding this public health issue, schools may play an important role in promoting physical fitness levels. Specifically, in most countries physical education (PE) teachers are required to carry out intervention programmes in order to improve students' health-related physical fitness levels (e.g. Ministerio de Educación y Ciencia, 2007; National Association for Sport and Physical Education, 2004).

However, a PE-based planning problem for developing students' health-related physical fitness levels is the necessity to "deliver" a large volume of curricular content during each academic year (Viciana et al., 2014). Additionally, PE potential is restricted by its limited curriculum time allocation (Hardman, 2008). Therefore, despite the fact that a long-term programme is the best way to improve physical fitness (Meyer et al., 2014), in many countries the application of a short-term physical fitness development programme is one feasible option in the PE setting (Viciana et al., 2014). In this line, previous PE-based studies found that a short-term intervention programme performed only twice a week can improve students' physical fitness levels (Faigenbaum and Mediate, 2006; Mayorga-Vega et al., 2013; Ramírez Lechuga et al., 2012; Santos et al., 2011; Weeks et al., 2008).

Another PE-based planning problem related to physical fitness is its expected decrease after a period of detraining. Previous studies have found that after eight to 12 weeks of detraining, children show a significant loss in the obtained gains in physical fitness levels (e.g. Ingle et al., 2006; Tsolakis et al., 2004). Therefore, previous authors have suggested that, after a physical fitness development programme, PE teachers should include some maintenance programmes to retain students' physical fitness levels throughout the entire academic year (Viciana et al., 2014). In this line, after a physical fitness development programme performed twice a week, Mayorga-Vega et al. (2013) carried out a maintenance programme once a week, while the students performed other PE curricular content in the other weekly session. This intervention programme permitted the retention of the physical fitness levels previously achieved; however, it involved consuming too much time available for PE lessons. Hence, this kind of maintenance programme would interfere with the normal development of other curricular content in PE.

A possible alternative for this problem could be the planning of physical fitness maintenance through other PE curricular content. This way, apart from maintaining the physical fitness levels previously obtained, these programmes would allow the teaching of other PE content at the same time. To our knowledge there are no previous studies examining the effect of a PE-based maintenance programme through other curricular content in order to maintain physical fitness levels achieved through a development programme. Consequently, the main purpose of the present study was to examine the effects of a PE-based development and maintenance programme on objective and perceived health-related physical fitness in high school students. A secondary purpose of this study was to analyze the objective and perceived physical activity levels during the PE-based development and maintenance programme.

Material and methods

Participants

The study protocol was first approved by the Ethical Committee of the University of Granada. Then, a public high school from an urban area in the Andalucía Region (Spain) was selected for the study. This high school was chosen because it had the appropriate conditions for the study: (a) an adequate indoor sports facility for the tests, and (b) a large number of classes per grade. After the school approvals were obtained, according to the school's suggestion all the students from the second grade (i.e. eighth grade of schooling) were invited to participate in the present study. Adolescents and their legal guardians were fully informed about all the features of the study and were required to sign an informed consent form.

For practical reasons and due to the nature of the present study (the intervention was focused on already established classes in a school setting) a cluster-randomized controlled trial was used (Mayorga-Vega et al., 2013; Merino-Marban et al., 2015). Therefore, the six established classes were assigned randomly to form one of the following study groups: control group (CG) or experimental group (EG). All the experimental and control sessions were performed by the same PE teacher of the participating school. The inclusion criterion was to have an attendance rate of

85% or higher for PE classes during the intervention period. Finally, a sample of 111 students, 70 boys and 41 girls, aged 12–14 years old agreed to participate and met the inclusion criterion. For general characteristics of the participants, see the Results section. According to the school reports, most of the students' families had a middle socioeconomic status. All the participants were free of any health disorder such as heart diseases, uncontrolled asthma, bone/ joint problems or other reasons why children should not undergo physical activity.

Measures

Evaluation was carried out during the PE classes at the beginning and at the end of the development intervention programme (pretest and posttest, respectively) in order to examine possible changes produced. Subsequently, after a period of detraining (coinciding with the Christmas holidays) and the application of the maintenance intervention programme, the participants were evaluated again in order to observe the levels of retention (retest). Each evaluation was carried out during two PE sessions of the same week. Students' perceived physical fitness and objective muscular fitness were assessed during the first test session. During the second test session, students' objective cardiorespiratory fitness was evaluated.

Each evaluation was carried out by the same tester, using the same instruments and under the same conditions. The measurements were taken in an indoor sports facility with a non-slippery floor, under the same environmental conditions, on the same day of the week and at the same time for each student. Prior to the objective physical fitness tests, the participants completed a standardized warm-up consisting of five minutes of running from low to moderate intensity followed by some joint mobility exercises. The order and a brief description of the measures protocol were as follows:

Perceived physical fitness. Perceived overall physical fitness, cardiorespiratory fitness and muscular fitness were measured using the three specific single-response items included in the International Fitness Scale (Ortega et al., 2011). The three 10-point Likert-scale items asked the participants to compare their perceived overall physical fitness, cardiorespiratory fitness and muscular fitness with their friends' physical fitness (from 1 = "very poor" to 10 = "very good"). Then, the students' perceived body image was also assessed by Thompson and Gray's (1995) Contour Drawing Rating Scale. For this purpose, participants were asked to choose among nine possible figures which one would best represent their own bodies. Both scales have demonstrated adequate reliability and validity among adolescents (Ortega et al., 2011; Viciana et al., 2015).

Objective muscular fitness. The flexion-extension legs test was used to assess the lower limbs' muscular endurance. Standing parallel close to the wall, the participants placed their feet at the width of their shoulders. Then, the participants raised their right upper-limb extended touching a mark with their fingers. From this position, the participants had to flex their knees until touching the floor with both their hands and then return to the starting position as many times as possible in 30 seconds. Participants were urged to fix their feet on the floor during the test, avoiding jumps or displacements. The participants carried out two repetitions before the test, and then they performed the test once. The total number of repetitions performed by each participant in 30 seconds was recorded. The flexion-extension legs test has shown adequate reliability and validity among adolescents (Martínez López, 2002).

Objective cardiorespiratory fitness. The 20-meter shuttle run test was used to assess cardiorespiratory fitness. The participants ran between two parallel lines placed 20 meters apart, in the rhythm marked by a recorded beep. The starting speed was 8.5 km/h and it increased 0.5 km/h every minute. The test ended when the participants stopped running because of fatigue or failed to reach the line before the next signal for two consecutive times. During the test each participant wore a heart rate monitor (Polar® RS300X, Finland). Participants were allowed to perform the test once. The total time in seconds was retained. The 20-meter shuttle run test has demonstrated adequate reliability and criterion-related validity among adolescents (Léger et al., 1988; Mayorga-Vega et al., 2015).

Procedures

A physical fitness programme was applied to the EG during the PE classes. In accordance with the established curriculum and the approval of the educational institution, the sessions were designed by the PE teacher with the supervision of the main researcher (DMV). Then, the intervention programme was carried out by the PE teacher. First, the EG students performed a development programme twice a week for nine weeks. Then, after a four-week period of detraining coinciding with the Christmas holiday, the EG participants completed a maintenance programme twice a week for eight weeks. Since two lessons of each programme coincided with holidays, in the end the EG students completed a total of 16 and 14 sessions of the development and maintenance programmes, respectively.

Each intervention session lasted approximately 50 minutes and consisted of a 5-to-10-minute warm-up, 35-to-40-minute main part, and five-minute cool-down. During the warm-up the students performed low-to-moderate aerobic activities followed by some joint mobility and stretching exercises of the main body parts. For instance, a general warm-up consisted of three to five minutes of jogging followed by six to eight joint mobility exercises (eight to 10 repetitions) and six to eight static stretching exercises (10–15 seconds) of the main regions. During the development programme, in the main part students performed commonly used PE-based physical fitness sessions (e.g. strength games, running games, circuit training, multi-jumps, or multi-throws) followed by some team games. For example, during the strength games session students performed six traditional strength games and their variants (tug-of-war, pushing game, piggy-back races, sumo fighting, and standing and all fours games that consist of throwing the opponent off balance) for five minutes each followed by the team game dodgeball for 10 minutes.

Regarding the maintenance programme, in the main part of the sessions students carried out 10–15 minutes of sports-integrated training activities followed by 20–30 minutes of introduction activities to these same sports (basketball, futsal, volleyball and acrosport). For example, the main part of a basketball session consisted of walking while bouncing a ball (two minutes), an endurance race while bouncing a ball (five minutes), and a three-minute circuit while bouncing a ball (two repetitions) for the sports-integrated training part; then, for the introduction activities to the sport, traditional basketball learning exercises were performed depending on the session objective (i.e. bounce, pass, throw, etc.); finally, students performed a match in a small-sided game situation such as 2×2, 3×3, and so on. During the whole intervention period one session a week was mainly focused on aerobic activities and the other one on muscular fitness activities. The PE teacher placed special emphasis on reaching a moderate-to-vigorous intensity during the experimental sessions. Finally, students carried out some conventional light activities during the cool-down period.

As regards the CG students, they also participated in their PE sessions twice a week during the intervention period. Similarly, the sessions lasted approximately 50 minutes and consisted of a

5-to-10-minute warm-up, 35-to-40-minute main part, and five-minute cool-down. However, the content and methodology followed during the main part of the sessions were different. During the development programme period, the CG performed sports sessions (volleyball and badminton) instead of physical fitness. On the other hand, during the maintenance programme period, the CG students were involved in similar sports to the EG students (basketball, hockey and acrosport). However, during both the development and maintenance programme, the EG sessions were based on the intensity of the tasks, implicating big muscle groups, and performing games. Meanwhile, the CG sessions were mostly based on technique-learning practice.

Intensity control. The intensity of all the sessions was controlled during the intervention period. At the end of each PE session all the students reported their global perceived exertion using a pictorial perceived exertion scale (from 0 = "not tired at all" to 10 = "very, very tired"). The pictorial perceived exertion scale has demonstrated adequate reliability and validity among adolescents (test-retest reliability, intraclass correlation coefficient (ICC) = 0.95; validity, $r = 0.89$) (Pfeiffer et al., 2002). Additionally, the heart rates of students were recorded during the sessions. For this purpose, in one session a week, five students from each class were randomly selected to wear a heart rate monitor (Polar® RS300X, Finland) during the whole session. One week prior to the first evaluation all the students performed a familiarization session with the pictorial perceived exertion scale during which they were taught how to use the scale. Then, they performed five to six activities of different intensities during which they had to self-report their perceived exertion. Then, in the next session of the same week, in addition to taking the anthropometric data, children were taught how to properly place the heart rate monitor.

Analysis

Besides some common exploratory analyses for potential data errors such as in the extreme cases (i.e. values greater than three standard deviations), additional examinations of data were carried out. For the 20-meter shuttle run test the scores of those participants who did not reach a heart rate value equal to or higher than 90% of estimated maximum heart rate were deleted. For the objective physical activity levels, the data of the participants whose heart rate had been registered less than 50% of the time of the session were also eliminated. Because of the deleted data for the above-mentioned exploratory analyses or simply because the participants did not attend some evaluation session, sample sizes vary slightly for the different variables. However, to make the maximum use of the data, all valid data on physical fitness and physical activity variables were included in this study.

Afterward, statistical tests assumptions (i.e. normality, homogeneity of variance, sphericity, etc.) were examined and met for all the tests conducted. Descriptive statistics (means and standard deviations) for general characteristics of the sample, objective and perceived physical fitness, and objective and perceived physical activity levels were calculated. Then, a one-way analysis of variance (ANOVA) was conducted to examine potential differences between the two groups in terms of body mass, height, body mass index, and baseline values of both objective and perceived physical fitness. Additionally, chi-squared analyses were carried out to test the ratio differences of gender and extracurricular sport practitioners between the two groups.

Subsequently, the effect of the PE-based development and maintenance programme on objective and perceived physical fitness was examined using a two-way ANOVA applied over the dependent variables (body mass index was used as a covariate when the objective cardiorespiratory fitness was analyzed), including *group* as an independent variable (CG, EG) and *time* as a dependent variable (pretest, posttest, retest). Afterwards, for the *post hoc* analyses, α values were corrected using the Bonferroni adjustment. Moreover, the Hedges' *g* effect size was used to examine the magnitude of treatment effects (Hedges, 2007). Additionally, the test-retest reliability of the dependent variables was estimated using the ICC from two-way ANOVA ($ICC_{3,\ k}$) (Shrout and Fleiss, 1979), as well as the 95% confidence interval. Finally, a one-way ANOVA was used to compare the objective and perceived physical activity levels during the PE-based development and maintenance programme between the EG and CG. All statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20). The statistical significance level was set at $p < 0.05$.

Results

General characteristics

The EG participants obtained an average attendance of 98% and 97% in the development and maintenance programmes, respectively. The general characteristics of the participants are shown in Table 1. The one-way ANOVA results did not show statistically significant differences in body mass, height, and body mass index values between the EG and CG ($p > 0.05$). Additionally, the chi-square analyses showed that the two groups had a balanced representation of boys and girls and extracurricular sport practitioners and non-practitioners ($p > 0.05$). Moreover, the one-way ANOVA results did not show statistically significant differences in the baseline values of any objective and perceived physical fitness variables between the EG and CG ($p > 0.05$).

Objective physical fitness

Table 2 shows the effect of the PE-based intervention programme on objective physical fitness levels. The results of the two-way analysis of covariance (ANCOVA) on the average obtained in the cardiorespiratory fitness test showed a significant interaction effect between the *group* and *time* variables ($F_{2,178} = 11.410$; $p < 0.001$). Subsequently, for *post hoc* analyses, the ANCOVA with the Bonferroni adjustment showed that the EG increased statistically significantly from pretest to posttest ($p = 0.032$) and from pretest to retest ($p = 0.001$). However, no significant differences were found from posttest to retest for the EG ($p = 0.098$). For the CG no significant differences were found ($p > 0.05$), except from the pretest to posttest, which decreased statistically significantly ($p = 0.034$). The test-retest reliability for cardiorespiratory fitness measure was 0.96 (0.93–0.98).

Regarding the muscular fitness test, the results of the two-way ANOVA showed a significant interaction effect between the *group* and *time* variables ($F_{2,196} = 5.139$; $p = 0.009$). Subsequently, for *post hoc* analyses, the ANOVA with the Bonferroni adjustment showed that the EG increased statistically significantly from pretest to posttest and from pretest to retest ($p < 0.001$). However, no significant differences were found from posttest to retest for the EG ($p = 0.723$). Regarding the CG, the ANOVA with the Bonferroni adjustment also showed statistically significant differences from pretest to posttest ($p = 0.001$) and from pretest to retest ($p = 0.047$), but not from posttest to retest ($p = 1.000$). The test-retest reliability for the muscular fitness measure was 0.92 (0.86–0.95).

Table 1. General characteristics (mean ± standard deviation/ frequency) of the participants and differences between experimental and control groups.

	Sample (n = 111)	Control (n = 57)	Experimental (n = 54)	p[a]
Age (year)	12.5 ± 0.7	12.6 ± 0.7	12.4 ± 0.6	–
Body mass (kg)	54.5 ± 16.0	56.9 ± 19.3	52.0 ± 11.2	0.106
Height (cm)	158.7 ± 7.8	159.9 ± 7.8	157.5 ± 7.6	0.117
Body mass index (kg/m^2)	21.4 ± 5.1	22.0 ± 6.2	20.8 ± 3.6	0.215
Gender (boys/ girls)	70 / 41	38 / 19	32 / 22	0.419
Extracurricular sport (yes/ no)[b]	86 / 25	46 / 11	40 / 14	0.403

[a]Significance level from the one-way analysis of variance for body mass, height and body mass index, and from the chi squared test for the gender and extracurricular sport ratios.
[b]Children that regularly participated (yes) or not (no) at least twice per week in extracurricular sport activities.

Table 2. Effect of the physical education-based intervention programme on objective physical fitness.

Group	Pretest (1) (M ± SD)	Posttest (2) (M ± SD)	Retest (3) (M ± SD)	p[a]	Effect size[b] 1–2	2–3	1–3
Cardiorespiratory fitness (s)							
Experimental (n = 50)	312.7 ± 128.0	327.7 ± 116.5[c]	338.9 ± 118.4[g]	< 0.001	0.25	0.08	0.33
Control (n = 42)	289.9 ± 119.5	273.7 ± 117.2[c]	275.3 ± 111.1				
Muscular fitness (n)							
Experimental (n = 52)	27.0 ± 3.5	29.2 ± 3.5[e]	29.5 ± 3.5[h]	0.009	0.29	0.16	0.44
Control (n = 48)	27.2 ± 3.4	28.4 ± 3.1[d]	28.2 ± 3.2[f]				

M: mean; SD: standard deviation.
[a]Significance level from two-way analysis of variance (and with the covariance body mass index for the variable cardiorespiratory fitness) with the *post hoc* analysis with Bonferroni adjustment: Change statistically significant from pretest to posttest ([c]p < 0.05, [d]p < 0.01, [e]p < 0.001) and from pretest to retest ([f]p < 0.05, [g]p < 0.01, [h]p < 0.01).
[b]Hedges' g effect size.

Perceived physical fitness

Table 3 shows the effect of the PE-based intervention programme on students' perceived physical fitness levels. The results of the two-way ANOVA on the average obtained in the perceived physical fitness values did not show significant interaction effects between the *group* and *time* variables: overall physical fitness ($F_{2,214} = 0.816$; $p = 0.437$), cardiorespiratory fitness ($F_{2,214} = 1.301$; $p = 0.273$), muscular fitness ($F_{2,214} = 0.784$; $p = 0.458$) and body image ($F_{2,214} = 0.995$; $p = 0.372$). The test-retest reliability for the perceived physical fitness values was good: overall physical fitness 0.87 (0.78–0.93), cardiorespiratory fitness 0.80 (0.67–0.89), muscular fitness 0.84 (0.72–0.90), and body image 0.85 (0.74–0.91).

Objective and perceived physical activity

Table 4 shows the comparison of the perceived and objective physical activity levels during the PE-based development and maintenance programme between the EG and CG. The results for the physical activity levels during the PE sessions in both the development programme and the

Table 3. Effect of the physical education-based intervention programme on perceived physical fitness.

Group	Pretest (1) ($M \pm SD$)	Posttest (2) ($M \pm SD$)	Retest (3) ($M \pm SD$)	p[a]	Effect size[b]		
					1–2	2–3	1–3
Overall physical fitness[c]							
Experimental ($n = 54$)	7.2 ± 1.6	7.0 ± 1.4	6.8 ± 1.6[f]	0.437	–0.03	–0.13	–0.15
Control ($n = 55$)	7.1 ± 1.5	6.9 ± 1.2	6.9 ± 1.4				
Cardiorespiratory fitness[c]							
Experimental ($n = 54$)	6.8 ± 1.8	6.7 ± 1.7	6.5 ± 1.8	0.237	0.24	–0.15	0.10
Control ($n = 55$)	6.8 ± 1.7	6.2 ± 1.6[e]	6.2 ± 1.6[f]				
Muscular fitness[c]							
Experimental ($n = 54$)	6.7 ± 1.8	6.6 ± 1.6	6.6 ± 1.7	0.458	0.16	–0.14	0.02
Control ($n = 55$)	6.9 ± 1.2	6.5 ± 1.4	6.8 ± 1.4				
Body image[d]							
Experimental ($n = 54$)	5.5 ± 1.2	5.3 ± 0.9	5.5 ± 1.1	0.372	–0.18	0.13	–0.07
Control ($n = 55$)	5.5 ± 1.2	5.5 ± 1.2	5.6 ± 1.1				

M: mean; SD: standard deviation.
[a]Significance level from two-way analysis of variance with the *post hoc* analysis with Bonferroni adjustment: Change statistically significant from pretest to posttest ([e]$p < 0.05$) and from pretest to retest ([f]$p < 0.05$).
[b]Hedges' g effect size.
[c]The scores ranged from 1 = "very poor" to 10 = "very good."
[d]The scores ranged from 1 to 9.

Table 4. Comparison of perceived and objective physical activity levels during the physical education-based development and maintenance programme between the experimental and control groups.

	Control		Experimental			
	n	($M \pm SD$)	n	($M \pm SD$)	p[a]	g[b]
Development programme						
RPE	56	3.0 ± 1.3	54	6.2 ± 1.1	< 0.001	2.71
Heart rate average (bpm)	50	131.0 ± 13.7	50	145.5 ± 14.4	< 0.001	1.04
Heart rate average (%)	50	64.5 ± 6.7	50	71.7 ± 7.1	< 0.001	1.04
MVPA (%)	50	35.6 ± 20.8	50	57.6 ± 18.9	< 0.001	1.11
Total time (minutes)	50	51.3 ± 2.2	50	50.5 ± 2.7	0.122	−0.31
Maintenance programme						
RPE	57	2.9 ± 1.2	54	5.7 ± 1.2	< 0.001	2.28
Heart rate average (bpm)	45	133.3 ± 16.5	44	147.4 ± 13.3	< 0.001	0.94
Heart rate average (%)	45	65.7 ± 8.1	44	72.6 ± 6.5	< 0.001	0.94
MVPA (%)	45	37.5 ± 25.3	44	59.3 ± 22.0	< 0.001	0.92
Total time (minutes)	45	49.6 ± 2.6	44	50.2 ± 1.7	0.212	0.27

M: mean; SD: standard deviation; RPE: rating of perceived exertion (the scores ranged from 0 = "not tired at all" to 10 = "very, very tired"); MVPA: moderate-to-vigorous physical activity (percentage of total time involved in an intensity ≥ 70% of maximum heart rate).
[a]Significance level from one-way analysis of variance.
[b]Hedges' g effect size.

maintenance programme were similar. The results of the one-way ANOVA found that the EG had statistically significantly greater levels of objective physical activity (i.e. average heart rate, average of the percentage of maximum heart rate, and percentage of total time involved in moderate-to-vigorous physical activity) than the CG ($p < 0.001$). Similarly, the EG participants reported statistically significantly higher values of perceived exertion rating than the CG ($p < 0.001$). On the other hand, no significant differences were found in the total time of PE sessions between groups ($p > 0.05$).

Discussion

Objective physical fitness

The main purpose of the present study was to examine the effects of a PE-based development and maintenance programme on objective health-related physical fitness in high school students. Planning a long-term physical fitness programme is the best way to improve physical fitness levels (Meyer et al., 2014). However, a PE-based planning problem for developing students' health-related physical fitness levels is the fact that much curricular content must be developed in a school year (e.g. health-related physical fitness, sports, body expression, or physical activities in the natural environment) (Viciana et al., 2014). Moreover, another planning-related problem is the fact that PE is restricted by its limited curriculum time allocation (Hardman, 2008). For instance, in several European countries PE is limited to only two sessions a week (European Commission/ EACEA/ Eurydice, 2013).

Consequently, since physical fitness programmes cannot last the whole academic year or a large part of it, in the PE setting the application of a short-term physical fitness development programme is one feasible option (Viciana et al., 2014). In this sense, the results of this study showed that a PE-based physical fitness programme performed twice a week for only nine weeks significantly improved both objective cardiorespiratory and muscular fitness in high school students. Similar to the present results, some previous PE-based studies found that a short-term intervention programme performed twice a week can significantly improve students' cardiorespiratory and muscular fitness (Faigenbaum and Mediate, 2006; Mayorga-Vega et al., 2013; Ramírez Lechuga et al., 2012; Santos et al., 2011; Weeks et al., 2008).

As regards the objective muscular fitness, the CG students also showed a statistically significant increase of their test scores compared to their baseline levels. However, these findings could be simply due to several previously described reasons such as maturation or testing effects (Thomas et al., 2011). Similar to the present study, previous short-term intervention studies have also found a statistically significant increase in CG students' scores (e.g. Weeks et al., 2008). Nevertheless, it must be highlighted that in the present study the two-way ANOVA results showed statistically significant interaction effects, and in the EG the intra-group comparisons showed a greater increase in the objective muscular fitness values. In this line, the effect size for the muscular fitness was also positive in favor of the EG. Therefore, the results of the present intervention showed a clear statistically significant positive effect on EG students' objective muscular fitness.

On the other hand, another PE planning-related problem is that physical fitness gains are expected to decrease after a period of detraining. PE teachers usually carry out a physical fitness programme for a few weeks, and then they teach other content without considering how long the effect will last. Furthermore, in addition to all the above-mentioned limitations that PE teachers have to face (i.e. many curricular contents must be developed in a school year and with limited curriculum time allocation), the academic year is frequently interrupted by several holiday periods. In this line, previous studies have found that after eight to 12 weeks of detraining children show a significant loss of the physical fitness obtained gains (e.g. Ingle et al., 2006; Tsolakis et al., 2004). Therefore, a maintenance programme should be applied in order to maintain the physical fitness levels previously gained during the rest of the academic year (called reinforced teaching units) (Viciana et al., 2014; Viciana and Mayorga-Vega, 2015).

Currently the evidence about the efficacy of physical fitness maintenance programmes among school-age children is still limited and contradictory (Blimkie et al., 1989; DeRenne et al., 1996), especially in the PE setting (Mayorga-Vega et al., 2013). On the one hand, DeRenne et al. (1996) found that a maintenance programme carried out once a week in young basketball players was efficient to retain muscular fitness. However, on the other hand, Blimkie et al. (1989) found that a maintenance programme carried out once a week in children was not sufficient to retain muscular fitness. Nevertheless, it should be noted that those previous studies were carried out in different settings and, also, in contrast to the present study, in those studies the maintenance programme was applied right after the development programme. Regrettably, as mentioned before, since academic periods are frequently alternated with holidays, this is not a real situation in a PE setting.

Similar to the present study, Mayorga-Vega et al. (2013) found that, after a physical fitness development programme carried out twice a week followed by a four-week period of detraining (coinciding with the Christmas holiday), a PE-based maintenance programme helped maintain the cardiorespiratory and muscular fitness levels previously obtained. This is the most common situation in PE planning because of the typical alternation of PE sessions with holiday periods. However, in the aforementioned study the maintenance programme was carried out once a week (alternated with sports content in the other weekly session) instead of through other curricular content such as in the present study. Although the Mayorga-Vega et al. (2013) intervention programme permitted the retention of the physical fitness levels previously achieved, by following this approach (i.e. alternating sports and fitness sessions during the maintenance programme) teaching health-related physical fitness content could consume too much of the time available in PE planning. Hence, this approach followed in the maintenance programme would interfere in the teaching of the other PE curricular content. Consequently, one of the most important outcomes of the present study was to find out that a maintenance programme delivered through sports activities allows the retention of physical fitness levels previously obtained. Moreover, a maintenance programme delivered through sports activities seems to be more feasible in a PE setting.

Regarding the magnitude of the effects of the intervention on objective physical fitness, previous studies with short-term PE-based physical fitness programmes carried out twice a week among high school students found similar results both for objective cardiorespiratory fitness (median $g = 0.21$, 0.16 to 0.31) (Ramírez Lechuga et al., 2012; Santos et al., 2011) and muscular fitness (median $g = 0.23$, –0.15 to 0.45) (Faigenbaum and Mediate, 2006; Santos et al., 2011; Weeks et al., 2008). Regarding the maintenance programme, however, to our knowledge there are no previous studies examining the effect of a PE-based maintenance programme throughout sports sessions. Nevertheless, similar to the present study, Mayorga-Vega et al. (2013) carried out a PE-based maintenance programme once a week while in the other session the schoolchildren performed sports sessions like the CG did. Although children improved muscular endurance slightly more than in the present study ($g = 0.27$ vs. 0.16), the cardiorespiratory fitness was not maintained like in the present study ($g = -0.11$ vs. 0.08). Therefore, the magnitude effects of the present study on the objective health-related physical fitness were moderate, indicating that the intervention was effective.

On the other hand, increasing training factors such as frequency, duration and/ or intensity of the intervention programme could have a positive consequence for the magnitude effects. In this line, Ramírez Lechuga et al. (2012) examined the effects of a PE-based programme and found that increasing the frequency of sessions from two to three times a week had a positive effect on cardiorespiratory fitness among adolescent girls ($g = 0.16$ vs. 0.56), but not among adolescent boys ($g = 0.18$ and 0.19, respectively). Similarly, Ardoy et al. (2011) found that, among high school students, simply doubling the PE sessions per week (i.e. from two to four sessions) had a significant increase in cardiorespiratory fitness within 16 weeks ($g = 0.40$). Moreover, these same authors found that, in addition to doubling the frequency of sessions per week (i.e. volume), when the intensity of the PE sessions was also increased the improvement was even higher ($g = 0.85$). Unfortunately, the preceding authors did not find the same increase in muscular strength ($g = –0.01/ –0.06$). Additionally, since in most European countries PE is limited only to two sessions per week (European Commission/ EACEA/ Eurydice, 2013), currently the application of these programmes seems to not be feasible in the PE setting.

Perceived physical fitness

The second main purpose of this study was to examine the effects of a PE-based development and maintenance programme on high school students' perceived physical fitness. Health is regarded not merely as the freedom from disease or injury, but also a state of complete physical and psychosocial well-being (World Health Organization, 1946). Physical self-concept, which is regarded as an important subdomain of overall self-concept, incorporates different characteristics such as perceived fitness and appearance (Marsh et al., 1994). In this line, perceived physical fitness attains relevance due to its potential effects on the levels of physical activity (Planinsec and Fosnaric, 2005), and in turn on physical and psychosocial well-being (Alfermann and Stoll, 2000).

Consequently, improving students' perceived physical fitness should also be an important issue in the PE setting. Unfortunately, the results of the present study showed that the PE-based development and maintenance programme did not show an influence on the students' perceived physical fitness variables studied. Previous studies examining the effects of PE-based physical fitness programmes on students' perceived physical fitness are limited (Mayorga-Vega et al., 2012; Sadres et al., 2001; Schmidt et al., 2013), especially among high school students (Schneider et al., 2008). Similar to the present study, Schneider et al. (2008) found that a PE-based intervention performed five sessions a week did not improve students' overall perceived physical fitness, or students' perceived cardiorespiratory fitness, muscular fitness and body image. Moreover, in most of the preceding studies carried out with elementary schoolchildren the intervention did not improve students' perceived physical fitness (Mayorga-Vega et al., 2012; Sadres et al., 2001; Schmidt et al., 2013; Schneider et al., 2008). On the contrary, Schmidt et al. (2013) found that a physical fitness intervention carried out twice a week for 10 weeks improved children's perceived cardiorespiratory fitness, but they did not find the same effects on students' perceived muscular fitness.

Regarding the magnitude effects of the intervention on students' perceived physical fitness, the present study showed similar results as previous studies for overall perceived physical fitness (median $g = –0.03$, –0.07 to 0.18) (Mayorga-Vega et al., 2012; Sadres et al., 2001; Schneider et al., 2008), perceived cardiorespiratory fitness (median $g = 0.18$, –0.10 to 0.21) (Mayorga-Vega et al., 2012; Schmidt et al., 2013; Schneider et al., 2008), perceived muscular fitness (median $g = 0.04$, 0.04 to 0.30) (Mayorga-Vega et al., 2012; Schmidt et al., 2013; Schneider et al., 2008) and body image (median $g = 0.10$, 0.05 to 0.14) (Mayorga-Vega et al., 2012; Schneider et al., 2008). Therefore, a physical fitness programme in children seems not to have the same positive influence on perceived physical fitness like in adults (e.g. Brazell-Roberts and Thomas, 1990). Apart from the physiological differences between children and adults, one reason for this difference could be the fact that in the present study the baseline values were high and, therefore, such scores could be difficult to increase as a result of a short-term programme (Mayorga-Vega et al., 2012).

Objective and perceived physical activity

A secondary purpose of this study was to examine students' objective and perceived physical activity levels during the PE-based development and maintenance programme. PE has been considered as a key opportunity for adolescents to contribute in the achievement of recommended moderate-to-vigorous physical activity levels for several reasons (Brusseau et al., 2011). Firstly, in many countries PE is a compulsory subject for all students. Secondly, PE has been shown to contribute significantly to adolescents' total daily physical activity (Brusseau et al., 2011). Ideally, children should spend at least 50% of their PE time on moderate-to-vigorous physical activity (United States Department of Health and Human Services, 2010). Several previous studies found that the above-mentioned target is rarely met in most PE sessions (Fairclough and Stratton, 2005). The results of the present study indicated that the EG students had greater levels of objective and perceived physical activity than the CG during the PE sessions both in development and maintenance programmes. Furthermore, another important outcome of the present study is the finding that the intervention programme allows the meeting of the above-mentioned target during PE sessions.

Recently, Lonsdale et al. (2013) carried out a systematic review of interventions designed to increase moderate-to-vigorous physical activity in PE lessons. Similar to the present study, most previous studies found that a PE-based intervention increased statistically significantly the time of moderate-to-vigorous physical activity in PE lessons. However, two of the seven studies included did not find a statistically significant increase in moderate-to-vigorous physical activity (Strand and Anderson, 1996; Verstraete et al., 2007). On the other hand, as regards perceived physical activity, previous studies examining the effect of a PE-based intervention on students' perceived physical activity levels were not found. Finally, regarding the magnitude effects of the intervention, the present study showed similar results to previous studies about interventions designed to increase moderate-to-vigorous physical activity (median $g = 0.84$, 0.16 to 2.79) (Lonsdale et al., 2013). Therefore, the magnitude effects of the present study were moderate, indicating that the intervention was effective in increasing health-enhancing physical activity levels during PE lessons.

Conclusions and future studies

In conclusion, to our knowledge this is the first study that examines the effect of a short-term physical fitness programme followed by a maintenance programme through sports activities in a PE setting. The results of the present study suggest that it is possible to develop and maintain objective cardiorespiratory and muscular fitness in the PE setting. However, this kind of intervention programme does not seem to change the students' perceived physical fitness variables studied. On the other hand, another important outcome of the present study is the fact that the intervention programme increased the objective and perceived physical activity levels, exceeding the recommended moderate-to-vigorous physical activity levels during PE sessions. In addition to the development and maintenance of physical fitness levels, at the same time the present intervention programme permits the regular development of other PE curricular content. Therefore, this finding could help teachers to design programmes that would permit a feasible and effective development and maintenance of health-related physical fitness in a PE setting. Future research studies should examine the effects of PE-based intervention programmes through other curricular content in order to maintain the physical fitness levels achieved through a development programme. Moreover, future research studies should also examine the effects of alternative PE-based approaches such as, for example, the Dynamic PE model in order to develop health-related physical fitness at the same time that students focus on other curricular content such as skills and games.

Acknowledgments

The authors gratefully acknowledge all the participating students and their parents, without whom the present study could not have been carried out. Additionally, the authors acknowledge all the members of the school for their enthusiasm and collaboration, especially the school principal and the physical education teachers. We also thank BSc Adrián Brenes Podadera for his great collaboration on the data collection. We thank Raquel Merino and Anna Szczesniak for the English revision. Finally, the authors would like to express our gratitude to the Editor Dr Nollaig McEvilly and two anonymous reviewers for all the hard work and detailed care that have undoubtedly improved the present manuscript.

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/ or publication of this article.

Funding

This work was supported by the Spanish Ministry of Education, Culture and Sport (grant number AP2010-5905 (Daniel Mayorga-Vega)).

References

Alfermann D and Stoll O (2000) Effects of physical exercise on self-concept and well-being. *International Journal of Sport Psychology* 31(1): 47–65.

Ardoy DN, Fernández-Rodríguez JM, Ruiz JR, et al. (2011) Improving physical fitness in adolescents through a school-based intervention: The EDUFIT study. *Revista Española de Cardiología* 64(6): 484–491.

Blair SN (2009) Physical inactivity: The biggest public health problem of the 21st century. *British Journal of Sports Medicine* 43(1): 1–2.

Blimkie CJ, Martin J, Ramsay J, et al. (1989) The effects of detraining and maintenance weight training on strength development in prepubertal boys. *Canadian Journal of Applied Sport Sciences* 14: S102.

Brazell-Roberts JV and Thomas LE (1990) Pumping iron and self-esteem: Starting a weight-training program increased the reported self-esteem of college-aged women. *Swimming Technique* 26: 21–23.

Brusseau TA, Kulinna PH, Tudor-Locke C, et al. (2011) Children's step counts on weekend, physical education, and non-physical education days. *Journal of Human Kinetics* 27: 125–135.

DeRenne C, Hetzler RK, Buxton BP, et al. (1996) Effects of training frequency on strength maintenance in pubescent baseball players. *Journal of Strength and Conditioning Research* 10(1): 8–14.

European Commission/ EACEA/ Eurydice (2013) *Physical education and sport at school in Europe Eurydice Report*. Luxembourg: Publications Office of the European Union.

Faigenbaum AD and Mediate P (2006) Effects of medicine ball training on fitness performance of high-school physical education students. *Physical Educator* 63(3): 160–167.

Fairclough S and Stratton G (2005) Physical activity levels in middle and high school physical education: A review. *Pediatric Exercise Science* 17(3): 217–236.

Hardman K (2008) Physical education in schools: A global perspective. *Kinesiology* 40(1): 5–28.

Hedges LV (2007) Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics* 32(4): 341–370.

Ingle L, Sleap M and Tolfrey K (2006) The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *Journal of Sports Sciences* 24(9): 987–997.

Kodama S, Saito K, Tanaka S, et al. (2009) Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. *JAMA* 301(19): 2024–2035.

Léger LA, Mercier D, Gadoury C, et al. (1988) The multistage 20 meter shuttle run test for aerobic fitness. *Journal of Sports Sciences* 6(2): 93–101.

Lonsdale C, Rosenkranz RR, Peralta LR, et al. (2013) A systematic review and meta-analysis of interventions designed to increase moderate-to-vigorous physical activity in school physical education lessons. *Preventive Medicine* 56(2): 152–161.

Marsh HW, Richards GE, Johnson S, et al. (1994) Physical Self-Description Questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology* 16(3): 270–305.

Martínez López EJ (2002) *Pruebas de aptitud física*. Barcelona: Paidotribo.

Mayorga-Vega D, Aguilar-Soto P and Viciana J (2015) Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *Journal of Sports Science and Medicine* 14(3): 536–547.

Mayorga-Vega D, Viciana J and Cocca A (2013) Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics* 37: 153–160.

Mayorga-Vega D, Viciana J, Cocca A, et al. (2012) Effect of a physical fitness program on physical self-concept and physical fitness elements in primary school students. *Perceptual and Motor Skills* 115(3): 984–996.

Merino-Marban R, Mayorga-Vega D, Fernandez-Rodriguez E, et al. (2015) Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *European Physical Education Review* 21(1): 83–92.

Meyer U, Schindler C, Zahner L, et al. (2014) Long-term effect of a school-based physical activity program (KISS) on fitness and adiposity in children: A cluster-randomized controlled trial. *PLoS One* 9(2): e87929.

Ministerio de Educación y Ciencia (2007) Real Decreto 1613/ 2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la educación secundaria obligatoria. *Boletín Oficial del Estado* 5: 677–773.

National Association for Sport and Physical Education (2004) *Moving into the future: National standards for physical education*. 2nd ed. Reston, VA: National Association for Sport and Physical Education.

Ortega FB, Ruiz JR, Castillo MJ, et al. (2005) Low level of physical fitness in Spanish adolescents. Relevance for future cardiovascular health (AVENA study). *Revista Española de Cardiología* 58(8): 898–909.

Ortega FB, Ruiz JR, Castillo MJ, et al. (2008) Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity* 32(1): 1–11.

Ortega FB, Ruiz JR, España-Romero V, et al. (2011) The International Fitness Scale (IFIS): Usefulness of self-reported fitness in youth. *International Journal of Epidemiology* 40(3): 701–711.

Pfeiffer KA, Pivarnik JM, Womack CJ, et al. (2002) Reliability and validity of the Borg and OMNI rating of perceived exertion scales in adolescent girls. *Medicine and Science in Sports and Exercise* 34(12): 2057–2061.

Planinsec J and Fosnaric S (2005) Relationship of perceived physical self-concept and physical activity level and sex among young children. *Perceptual and Motor Skills* 100(2): 349–353.

Ramírez Lechuga J, Muros Molina JJ, Morente Sánchez J, et al. (2012) Efecto de un programa de entrenamiento aeróbico de 8 semanas durante las clases de educación física en adolescentes. *Nutrición Hospitalaria* 27(3): 747–754.

Ruiz JR, Castro-Pinero J, Artero EG, et al. (2009) Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine* 43(12): 909–923.

Sadres E, Eliakim A, Constantini N, et al. (2001) The effect of long-term resistance training on anthropometric measures, muscle strength, and self-concept in pre-pubertal boys. *Pediatric Exercise Science* 13(4): 357–372.

Santos A, Marinho DA, Costa AM, et al. (2011) The effects of concurrent resistance and endurance training follow a specific detraining cycle in young school girls. *Journal of Human Kinetics* Special Issue: 93–103.

Schmidt M, Valkanover S, Roebers C, et al. (2013) Promoting a functional physical self-concept in physical education: Evaluation of a 10-week intervention. *European Physical Education Review* 19(2): 232–255.

Schneider M, Dunton GF and Cooper DA (2008) Physical activity and physical self-concept among sedentary adolescent females: An intervention study. *Psychology of Sport and Exercise* 9(1): 1–14.

Shrout PE and Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2): 420–428.

Strand B and Anderson C (1996) A comparison of two instructional formats on heart rate intensity, skill achievement and student attitude. *Nebraska Journal for the Nebraska Association of Health, Physical Education and Recreation* 27(2): 21–24.

Thomas JR, Nelson JK and Silverman SJ (2011) *Research methods in physical activity*. 6th ed. Champaign, IL, USA: Human Kinetics.

Thompson MA and Gray JJ (1995) Development and validity of a new body-image assessment scale. *Journal of Personality Assessment* 64(2): 258–269.

Tsolakis CK, Vagenas GK and Dessypris AG (2004) Strength adaptations and hormonal responses to resistance training and detraining in preadolescent males. *Journal of Strength and Conditioning Research* 18(3): 625–629.

United States Department of Health and Human Services (2010) *Strategies to improve the quality of physical education*. Washington: Centers for Disease Control and Prevention.

Verstraete SJM, Cardon GM, De Clercq DLR, et al. (2007) Effectiveness of a two-year health-related physical education intervention in elementary schools. *Journal of Teaching in Physical Education* 26(1): 20–34.

Viciana J and Mayorga-Vega D (2015) Innovative teaching units applied to physical education. Changing the curriculum management for authentic outcomes. *Kinesiology*. Epub ahead of print 15 April 2015.

Viciana J, Mayorga-Vega D and Martínez-Baena AC (2015) La escala de imagen corporal de Thompson y Gray (1995) es también fiable y válida entre los adolescentes. *Revista Andaluza de Medicina del Deporte* 8(1): 46.

Viciana J, Mayorga-Vega D and Merino-Marban R (2014) Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In: Todaro R (ed) *Handbook of physical education research. Role of school programs, children's attitudes and health implications*. New York: Nova Science Publisher, pp.237–252.

Weeks BK, Young CM and Beck BR (2008) Eight months of regular in-school jumping improves indices of bone strength in adolescent boys and girls: The POWER PE study. *Journal of Bone and Mineral Research* 23(7): 1002–1011.

World Health Organization (1946) *Constitution of the World Health Organization*. New York: International Health Conference.

Author biographies

Daniel Mayorga-Vega is a Research fellow in the Department of Physical Education and Sport, University of Granada, Spain.

Jorge Montoro-Escaño is a Teacher in the IES Torre Atalaya, Junta de Andalucía, Spain.

Rafael Merino-Marban is a Professor in the Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Spain.

Jesús Viciana is a Professor in the Department of Physical Education and Sport, University of Granada, Spain.

Paper XVI

EFFECTS OF A STRETCHING DEVELOPMENT AND MAINTENANCE PROGRAM ON HAMSTRING EXTENSIBILITY IN SCHOOLCHILDREN: A CLUSTER-RANDOMIZED CONTROLLED TRIAL

Mayorga-Vega, D., Merino-Marban, R., Manzano-Lagunas, J., Blanco, H., & Viciana, J.

Journal of Sports Science and Medicine

Pending acceptance of the revised version

EFFECTS OF A STRETCHING DEVELOPMENT AND MAINTENANCE PROGRAM ON HAMSTRING EXTENSIBILITY IN SCHOOLCHILDREN: A CLUSTER-RANDOMIZED CONTROLLED TRIAL

Mayorga-Vega, D.[1], Merino-Marban, R.[2], Manzano-Lagunas, J.[3], Blanco, H.[4] and Viciana, J.[1]

[1] Department of Physical Education and Sport, University of Granada, Spain.

[2] Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Spain.

[3] CEIP Virgen de Belén, Junta de Andalucía, Spain.

[4] Faculty of Physical Education and Sport Sciences, University of Chihuahua, Mexico.

Running head: Stretching development and maintenance program

Abstract

The main purpose of the present study was to examine the effects of a physical education-based stretching development and maintenance program on hamstring extensibility in schoolchildren. A sample of 150 schoolchildren aged 7-10 years old from a primary school participated in the present study (140 participants were finally included). The six classes balanced by grade were cluster randomly assigned to the experimental group 1 (n = 51), experimental group 2 (n = 51) or control group (n = 49) (i.e., a *cluster randomized controlled trial* design was used). During the physical education classes, the students from the experimental groups 1 and 2 performed a four-minute stretching program twice a week for nine weeks (first semester). Then, after a five-week period of detraining coinciding with the Christmas holidays, the students from the experimental groups 1 and 2 completed another stretching program twice a week for eleven weeks (second semester). The students from the experimental group 1 continued performing the stretching program for four minutes while those from the experimental group 2 completed a flexibility maintenance program for only one minute. The results of the two-way analysis of variance showed that the physical education-based stretching development program significantly improved the students' hamstring extensibility ($p < 0.001$), as well as that these gains obtained remained after the stretching maintenance program ($p < 0.001$). Additionally, statistically significant differences between the two experimental groups were not found ($p > 0.05$). After a short-term stretching development program, a physical education-based stretching maintenance program of only one-minute sessions twice a week is effective in maintaining hamstring extensibility among schoolchildren. This knowledge could help

and guide teachers to design programs that allow a feasible and effective development and maintenance of students' flexibility in the physical education setting.

Key words: Flexibility program, flexibility maintaining, classic sit-and-reach test, physical fitness, primary school children, physical education setting.

Introduction

Physical fitness is considered one of the most important health markers in childhood (Ortega et al., 2008), with flexibility being one of the key components of health-related physical fitness (Meredith and Welk, 2010). For instance, the lack of hamstring extensibility creates a decrease in pelvic mobility (López-Miñarro et al., 2012). Therefore, when individuals with poor hamstring extensibility perform a maximal trunk flexion with straight legs, a posterior pelvic tilt and an increase in spinal flexion occur (López-Miñarro et al., 2012). This invariably leads to biomechanical changes in the pressure distribution of the spine, which may condition spinal disorders (Carregaro and Coury, 2009). Particularly among children, low hamstring extensibility contributes to an increase in the risk of current low back pain (Feldman et al., 2001; Jones et al., 2005; Sjölie, 2004) and neck tension (Mikkelsson et al., 2006), as well as a higher risk of low back pain later in adulthood (Kujala et al., 1994).

Unfortunately, nowadays about one in five schoolchildren have a hamstring extensibility level indicative of health risk (Castro-Piñero et al., 2013). Therefore, health promotion policies should also be designed to identify schoolchildren with low hamstring extensibility as well as to encourage them to achieve healthy levels of extensibility. Regarding this public health issue, schools may play an important role in promoting health-enhancing flexibility levels (Ortega et al., 2008). Specifically, shortened hamstring muscles could be addressed proactively by systematically performing stretching exercises during physical education (PE) classes (Santonja Medina et al., 2007; Thacker et al., 2004). Current trends in most countries include requiring PE teachers to achieve and maintain schoolchildren's health-enhancing flexibility levels (e.g., Ministerio de Educación y Ciencia, 2006; National Association for Sport and Physical Education, 2005).

Previous studies have found that a PE-based stretching program carried out twice a week improves hamstring extensibility in schoolchildren (e.g., Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015; Sanchez Rivas et al., 2014). However, current PE teachers face several planning-related problems for developing students' flexibility levels (Viciana et al., 2014). For instance, many curricular contents must be developed each academic year. Therefore, stretching exercises cannot be allocated a large part of the time available in PE planning. Additionally, PE is usually limited by its restricted curriculum time allocation (Hardman, 2008), particularly when the weekly frequency of sessions is only twice a week, which is the norm in most European countries (European Commission/ EACEA/ Eurydice, 2013).

Another PE-based planning problem related to flexibility is its expected decrease after a period of detraining (Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015). On the one hand, besides the large volume of curricular contents in relation with the restricted curriculum time allocation, a limitation for PE-based planning is the fact that the academic year is frequently interrupted by several holiday periods such as Christmas holidays or Holy week (Viciana et al., 2014). On the other hand, many PE teachers conceive planning as "watertight drawers" that they have to fill with curricular contents (Siedentop and Tanehill, 2000). Therefore, PE teachers usually carry out stretching exercises in their classes only for some weeks, and when they cease doing them, they do not concern themselves with how long the effect will last (Merino-Marban et al., 2015). In this line, Viciana et al. (2014) suggested that after a stretching development program PE teachers should include a maintenance program in order to retain students' flexibility levels throughout the whole academic year. Apart from maintaining the flexibility levels previously obtained, these programs would not interfere in the normal teaching of other PE curricular contents (Viciana et al., 2014).

Unfortunately, to our knowledge there are no previous studies examining the effect of a PE-based stretching maintenance program among schoolchildren. Consequently, the main purpose of the present study was to examine the effects of a PE-based stretching development and maintenance program on hamstring extensibility in schoolchildren. A secondary purpose of this study was to compare the effect of the PE-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline. It was hypothesized that a PE-based stretching development program would improve schoolchildren´s hamstring extensibility, as well as that a one-minute maintenance program would retain the flexibility gains obtained previously. Finally, it was also hypothesized that children with low hamstring extensibility would improve more than those with normal hamstring extensibility.

Methods

Participants

The study protocol was first approved by the Ethical Committee of the University of Granada. After the school approvals were obtained, schoolchildren and their legal guardians were fully informed about all the features of the study [i.e., a thorough description of the methods, potential risks, expected benefits, etc.; based on Thomas´s et al. (2015) guidelines] and were required to sign an informed consent document. A sample of 150 schoolchildren, 70 boys and 80 girls, aged 7-10 years old from six different third/ fourth-grade PE classes of a public primary school center participated in the present study. For practical reasons and the nature of the present study (i.e., intervention focused on natural groups in a school setting) a cluster randomized controlled trial design was used (Mayorga-Vega et al., 2013; Merino-Marban et al., 2015). The six natural classes balanced by grade were assigned randomly to form one of the following study groups: experimental group 1 (EG1), experimental group 2 (EG2) or the control group (CG).

All the participants were free of orthopedic disorders such as episodes of hamstring and/ or lumbar injuries, fractures, surgery or pain in the spine or hamstring and/ or lumbar muscles over the past six months. The inclusion criteria were: (a) correctly performing all the flexibility evaluations, and (b) having an attendance rate of 90% or higher for PE classes during the intervention period. Finally, although all the 150 invited schoolchildren agreed to participate, only 140 participants met the inclusion criteria. Figure 1 shows the flow chart that corresponds with the participants included in the present study. For general characteristics of the included participants, see the Results section.

Measures

Evaluation was carried out during the PE classes at the beginning and at the end of the stretching development program (pre-intervention and post-development, respectively) in order to examine possible changes produced. Subsequently, after a period of detraining (coinciding with the Christmas holidays) and the application of the stretching maintenance program, the participants were evaluated again in order to observe the levels of retention (post-maintenance). Each evaluation was carried out by the same evaluator, instrument, and under the same conditions. All the measures were taken in an indoor sports facility on the same day of the week and at the same time for each participant. Because of practical reasons, no warm-up exercises were performed prior to the test. The participants were assessed in sportswear and barefeet.

The classic sit-and-reach (SR) test was used to estimate participants' hamstring extensibility. Briefly, at the beginning of the test the participants stood in front of the box, sat with their hips flexed, knees extended and both hands on the top of the ruler. From this position the participants had to bend the trunk forward slowly and progressively (no swings) in order to reach the furthest possible distance and to remain still for at least two seconds (a score of 16 cm corresponded to the tangent of the feet; accuracy ± 0.5 cm). The participants were allowed to perform the test twice with one minute apart and then the average score in cm was retained (Mayorga-Vega et al., 2015). Additionally, participants' hamstring extensibility was categorized as follows: $<$ 14.0 cm low and $\geq$ 14.0 normal hamstring extensibility (Ferrer, 1998). The SR test has demonstrated high reliability (ICC = 0.99) (Ayala et al., 2012) and adequate criterion-related validity (r_p = 0.67, 0.55-0.79) among children (Mayorga-Vega et al., 2014c).

Procedures

A stretching intervention program was applied to the EGs during the cool-down period of their PE sessions (Mayorga-Vega et al., 2014a). Initially, the EG students performed a four-minute stretching development program twice a week for nine weeks (first semester). Then, after a five-week period of detraining coinciding with the Christmas holidays, the EG participants completed another stretching intervention program twice a week for eleven weeks (second semester). The EG1 students continued performing the stretching intervention program for four minutes while the EG2 students completed a maintenance program for only one minute. Since two lessons of the second semester coincided with holidays, in the end the EG participants completed a total of 18 and 20 sessions of the development and maintenance programs, respectively. In the PE setting the intervention performed with the EG2 has been called "reinforced teaching unit" (Viciana and Mayorga-Vega, 2015).

During each intervention session, the EG participants performed hamstring stretches using the static technique. Each intervention session included two 30-second sets of four stretching exercises (except for the maintenance sessions of the EG2 where only one stretching exercise was included). Six different stretching exercises were designed and alternated during the intervention program (Figure 2). Three bipodal exercises and one unipodal exercise were performed in each session (except for the maintenance sessions of the EG2 that only an exercise was performed). In all the stretching exercises, the students were placed with their hips flexed and knees extended. From this position the students flexed forward at the hip trying to maintain the spine in a neutral position as much as possible until a gentle stretch was felt in the hamstrings. The stretched position was held gently until the end point of the range of motion was reached (i.e., stretch to the point of feeling the tightness, but no pain). Once this position was achieved, the participants held it for 30 seconds. During the first two sessions of the intervention the PE teacher explained in depth how to properly perform the stretching exercises. Additionally, students also received constant feedback on their execution every session during the whole intervention. On the other hand, since teaching how to properly develop flexibility is a mandatory objective for PE in Spain (Ministerio de Educación y Ciencia, 2006), schoolchildren learn how to perform stretching exercises from the first grade of schooling (i.e., 6 years old).

All the students were urged to maintain their normal levels of physical activity outside the supervised setting during the research period. During the intervention program period all the students participated in their standard PE sessions. However, the CG students did not perform stretching exercises and were not aware of the purpose of the study. Both the standard PE sessions and the stretching intervention programs were carried out by the same PE teacher of the participating center for all the groups.

Statistical analysis

Descriptive statistics (means ± standard deviations or frequency) for age, gender, body mass, body height, body mass index, extracurricular sport, hamstring extensibility, and SR scores were calculated. A one-way analysis of variance (ANOVA) was conducted to examine potential differences between the three groups in terms of body mass, body height, body mass index, and baselines values of the SR test. Additionally, a chi-squared analysis was carried out to test the ratio differences of gender, extracurricular sport and pre-intervention hamstring extensibility categories between the three groups. The reliability of the SR scores was estimated using the intraclass correlation coefficient from the two-way ANOVA (Shrout and Fleiss, 1979), as well as the 95% confidence interval. Afterwards, the effect of the PE-based development and maintenance program on hamstring extensibility was examined using a two-way ANOVA applied over the SR scores, including *group* as an independent variable (EG1, EG2, CG) and *time* as a dependent variable (pre-intervention, post-development, post-maintenance). Subsequently, the *post hoc* analyses with the Bonferroni adjustment was used for both between or within-groups pairwise comparisons. Moreover, the Hedges' *g* effect size was used to examine the magnitude of the intervention effects (Hedges, 2007). The minimal detectable change was calculated in order to examine if the change score due to the intervention was true and reliable rather than the measurement error (Haley and Fragala-Pinkham, 2006).

On the other hand, the exacted McNemar's test was calculated in order to examine if the PE-based stretching development and maintenance program increased the proportion of children with normal hamstring extensibility. Finally, the two-way ANOVA, Hedges' *g* effect size, and minimal detectable change were also calculated in order to examine the effect of the PE-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline. For this last purpose, since most of the schoolchildren with low hamstring extensibility moved to a normal hamstring extensibility level after the development program and the total experimental sample was divided into two different interventions during the maintenance program period, data from both EGs were combined and only the effect of the development program was examined. All statistical analyses were performed using the SPSS version 20.0 for Windows (IBM® SPSS® Statistics 20). The statistical significance level was set at $p < 0.05$.

Results

Table 1 shows the general characteristics of the participants studied. The one-way ANOVA results did not show statistically significant differences in body mass, body height, body mass index, and SR baseline values between groups ($p > 0.05$). Furthermore, the chi-square analysis showed that the three groups had a balanced representation of boys/ girls, extracurricular sport practitioners/ non-practitioners, and low/ normal hamstring extensibility ($p > 0.05$). The reliability of the SR scores was 0.97 (0.96-0.98).

Table 2 shows the effect of the PE-based stretching intervention program on hamstring extensibility levels. The results of the two-way ANOVA on the average obtained in the SR test showed a statistically significant interaction effect between the *group* and *time* variables [$F(4, 274) = 6.177$; $p < 0.001$; $\eta^2 = 0.083$; $P = 0.987$]. Subsequently, the between-group pairwise comparisons with the Bonferroni adjustment showed that both the EG1 and EG2 were statistically significantly different than the CG ($p < 0.05$). Nevertheless, statistically significant differences between the EG1 and EG2 were not found ($p > 0.05$). Regarding the within-group analyses, the pairwise comparisons with the Bonferroni adjustment showed that both the EG1 and EG2 statistically significantly improved hamstring extensibility from pre-intervention to post-development ($p < 0.001$) and from pre-intervention to post-maintenance ($p < 0.001$). However, for both the EG1 and EG2 from post-development to post-maintenance no statistically significant differences were found ($p > 0.05$). Additionally, statistically significant differences were found for the CG ($p > 0.05$). The minimal detectable change value of the SR score was 2.2 cm, when the average increase in the EG1 was 2.7 and 3.3 cm and in the EG2 2.4 and 2.3 cm for the pre-intervention-post-development and pre-intervention-post-maintenance, respectively.

Table 3 shows the effect of the PE-based stretching intervention program on the proportion of children with normal hamstring extensibility. The results of the exact McNemar's test showed that for both the EG1 and EG2 there was a statistically significant increase on the proportion of children with normal hamstring extensibility from pre-intervention to post-development ($p < 0.05$). Additionally, for both the EG1 and EG2 from post-development to post-maintenance no statistically significant differences were found ($p > 0.05$). On the other hand, statistically significant differences were not found for the CG ($p > 0.05$).

Figure 3 shows the effect of the PE-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline. The results of the two-way ANOVA on the average obtained in the SR test showed a statistically significant interaction effect between the *group* and *time* variables [$F(2, 137) = 12.998$; $p < 0.001$; $\eta^2 = 0.159$; $P = 0.997$]. Subsequently, the between-group pairwise comparisons with the Bonferroni adjustment showed that both the low and normal groups were different than the CG in a statistically significant manner ($p < 0.05$; g = 0.70 and 0.34, respectively). Additionally, statistically significant differences between the low and normal groups were found in favor of students with low flexibility baseline levels ($p < 0.001$; $g = 0.36$). Regarding the within-group analyses, the pairwise comparisons with the Bonferroni adjustment showed that both the low and normal groups had a statistically significant improvement in hamstring extensibility from pre-intervention to post-development ($p < 0.001$). However, statistically significant differences were not found for the CG ($p > 0.05$). Although the average increase in the children with low hamstring extensibility (3.9 cm) was above the minimal detectable change value (i.e., 2.2 cm), in the students with normal hamstring extensibility it was slightly below (1.9 cm).

Discussion

The main purpose of this study was to examine the effects of a PE-based stretching development and maintenance program on hamstring extensibility in schoolchildren. Similarly to previous studies carried out only twice a week (e.g., Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015; Sanchez Rivas et al., 2014), the findings of the present study have shown that a PE-based stretching development program improves hamstring extensibility in schoolchildren. Furthermore, these outcomes also showed that the short-term stretching development program increased the proportion of schoolchildren with normal hamstring extensibility. Regarding the magnitude effects of the development program, in the present study the effect size of the stretching development program was moderate, indicating that the intervention was effective. Similarly to the present results, all the previous studies carrying out short-term PE-based stretching programs obtained on average similar effect sizes (g = 0.43) (Kamandulis et al., 2013; Mayorga-Vega et al., 2014a; Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015; Sánchez Rivas et al., 2014).

Increasing training factors such as the frequency or duration of the stretching intervention program could have a positive consequence on the magnitude effects. In regard to the frequency of the stretching program, Santonja et al. (2007) found that when the children performed four sessions per week instead of two, the magnitude effect doubled (twice a week, g = 0.85; four times a week, g = 1.53; both 5 minutes per session, 31 weeks of duration). However, since in most European countries PE is limited to only two sessions per week (European Commission/ EACEA/ Eurydice, 2013), the application of stretching programs with a high frequency is not feasible in this educational setting. Merino-Marban et al. (2015) indicated that the increase of active time for learning in extra-curricular periods would represent an excellent strategy for PE teachers to pursue important objectives such as the flexibility improvement. Nevertheless, since this strategy mainly depends on the students' autonomy, we have to be aware that in primary schoolchildren it could be impractical (Merino-Marban et al., 2015).

In regard to the duration of the stretching program, while the previous studies examining the short-term stretching programs effects (5-10 weeks) obtained similar effect sizes as the current study (g = 0.43, 0.24-0.67) (Kamandulis et al., 2013; Mayorga-Vega et al., 2014a; Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015;

Sánchez Rivas et al., 2014), the magnitude was higher for the mid-term stretching programs (16 weeks) (*g* = 0.86, 0.85-0.88) (Coledam et al., 2012), and even higher for those with long-term stretching programs (whole school year, 31-32 weeks) (*g* = 0.94, 0.85-2.06) (Rodríguez et al., 2008; Sainz de Baranda 2009; Sainz de Baranda et al., 2006; Santonja Medina et al., 2007). However, in most European countries the application of a stretching development program with a relatively high time per session during the whole academic year is not suitable in the PE setting (Viciana et al., 2014).

Since stretching programs cannot be allocated a large part of PE time, Viciana et al. (2014) suggested that PE teachers should include a maintenance program with a reduced volume in order to retain students' flexibility levels gained during the previous semesters. The results of the present study showed that a PE-based stretching maintenance program carried out only one minute per session (i.e., the 25% volume of stretching per session than in the previous development program) retains schoolchildren's hamstring extensibility gains previously obtained. Additionally, the results of this study showed that the one-minute maintenance program was not different comparing to the four-minute maintenance program (i.e., when the EG students continued performing the same volume of stretching that in the previous development program). Similarly, the findings of this study also showed that the one-minute maintenance program retained the number of schoolchildren with normal hamstring extensibility as the four-minute maintenance program did. However, as regards the effect size, although the maintenance program carried out for one minute per session was also moderate, the magnitude of the EG participants that were enrolled in sessions for four minutes was slightly higher.

Regarding the previous literature, studies examining the effect of a stretching maintenance program in schoolchildren were not found, and the number of related studies with adults was scarce as well. On the one hand, Rancour et al. (2009) found that, after a daily development stretching program, with 2-3 sessions per week the adults also maintained the flexibility levels previously gained. On the other hand, Willy et al. (2001) carried out a study with adults that stretched five days per week for six weeks followed by four weeks of cessation and then they carried out six weeks of resumption following the same protocol as in the development program. Similar to the results of the current study with the EG1, although after the resumption program adults obtained statistically significantly gains, the scores obtained were similar that at the end of the development program.

A secondary purpose of the present study was to compare the effect of the PE-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline. The results of this study showed that both schoolchildren with low and normal flexibility baseline levels improved their hamstring extensibility. However, this improvement was higher among the children with low hamstring extensibility than those with normal flexibility baseline levels. Unfortunately, to our knowledge there are no previous studies examining the influence of the children´s flexibility baseline levels on the effect of a PE-based stretching intervention program. Similarly to the present study, Mayorga-Vega and Viciana (2015) examined the effects of a PE-based physical fitness program on cardiorespiratory fitness according to the students´ baseline levels. Although previous authors also found that the students with a lower physical fitness baseline improved their cardiorespiratory fitness levels, no significant differences between the students with higher levels and control students were found. In this line, despite the fact that in the present study significant differences were found for both experimental groups against the CG, only the increase in the children with low hamstring extensibility was above the minimal detectable change value. This points out that the improvement score observed in the students with normal hamstring extensibility could be not true and reliable rather than the measurement error (Haley and Fragala-Pinkham, 2006). Moreover, meanwhile the effect size of the stretching development program for schoolchildren with low hamstring extensibility was high, for their classmates with normal baseline flexibility levels the improvement was low.

Finally, as regards the physiological mechanics responsible for the increase in hamstring extensibility found, some theories have been proposed. Traditionally, most of the theories suggest that the increases in muscle extensibility observed after a stretching program involve a mechanical increase in length of the stretched muscle. These theories, which are known as “Mechanical theories”, include viscoelastic deformation, plastic deformation, increased sarcomeres in series, and neuromuscular relaxation. Nevertheless, up to date evidence does not support any of these theories. Firstly, although some experimental studies found that muscle length does increase during stretch application due to the viscoelastic deformation of muscle, all of these studies consistently showed viscoelastic deformation to be transient. Then, there is simply no empirical evidence of a plastic deformation phase in muscles. Regarding the increased sarcomeres in series, studies with animals have demonstrated that when muscles are immobilized in a fully extended position, there is an increase in the number of sarcomeres in series. However, these muscles do not undergo overall change in muscle length because it is offset by a concurrent decrease in sarcomere length. Finally, it has also often been suggested that involuntary contraction of muscles due to a neuromuscular “stretch reflex” can limit muscle elongation during stretching and, therefore, individuals should stretch slowly in order to stimulate neuromuscular reflexes that induce relaxation of muscles. Nevertheless, no study has supported the above mentioned assertions.

More recently, a new theory, which is referred as the “Sensory theory”, suggests that increases in muscle extensibility are only due to a modification of individual´s sensation and not to an increase in muscle length. Most of the studies support the notion that increases in muscle extensibility observed after a single stretching session and after short-term stretching programs are due to a modification of the individual´s sensation. However, the effects of mid-term and long-term stretching programs have not yet been examined. Therefore, in the present study the increases in hamstring extensibility observed after the short-term stretching development program could be explained only by modifications in children’s sensation. Likewise, the Sensory theory may also explain the retention in hamstring extensibility during the stretching maintenance program. A more extensive explanation of physiological mechanisms of improvements in muscle extensibility due to a stretching program can be found elsewhere (e.g., Gajdosik, 2001; Weppler and Magnusson, 2010).

Limitations and future research studies

One of the most important limitations of the present study was to not perform a warm-up during the flexibility evaluation sessions. Although it has been observed that performing a warm-up increases the flexibility test scores (Díaz-Soler et al., 2015; O´Sullivan et al., 2009; Ryan et al., 2014), in the present study no warm-up exercises were performed. Firstly, previous studies have shown that the acute effect of a warm-up only lasts some minutes (DePino et al., 2000; O´Sullivan et al., 2009; Spernoga et al.

2001). Secondly, the effects differ significantly if the test is performed immediately after the warm-up or after a certain time (DePino et al., 2000; Díaz-Soler et al., 2015; O´Sullivan et al., 2009; Spernoga et al. 2001). However, because of practical reasons, the test would be only executed after a warm-up by all students simultaneously, but because of the order of assessment, some students would take the test immediately after finishing the warm-up and others would do it after several minutes. Therefore, in order to ensure the same conditions for all the participants (i.e., avoiding extraneous variables), no warm-up exercises were performed prior to the SR test.

On the other hand, since in the present study the pre-intervention values were assessed, any change in the participants' flexibility was compared with their baseline levels. Therefore, it could reasonably assume that any change in participants´ hamstring extensibility due to the stretching intervention compared to their pre-intervention values would be the same after the application or not of a warm-up protocol (i.e., following the same protocol in each flexibility evaluation session). In this line, similarly to the present study, none of the previous studies examining the effect of stretching intervention programs in schoolchildren performed warm-up exercises prior to the flexibility evaluation (e.g., Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015; Rodríguez et al., 2008; Sainz de Baranda 2009; Santonja Medina et al., 2007). Furthermore, since in the numerous previous studies no injury was reported as in the present study, not performing a warm-up prior to the flexibility evaluation is demonstrated to be completely safe among children.

Another important limitation in the present study was the fact that the post-detraining values were not measured. Although hamstring extensibility improvements are expected to decrease after a period of detraining (Mayorga-Vega et al., 2014b;

Merino-Marban et al., 2015), the PE-based stretching programs are frequently interrupted by several holiday periods (Viciana et al., 2014). Regrettably, in the current study the detraining effect during the Christmas holidays could not be examined for practical reasons. However, previous studies where a PE-based stretching program was carried out observed statistically significant loss of hamstring extensibility after five weeks of detraining (Mayorga-Vega et al., 2014b; Merino-Marban et al., 2015). Therefore, PE teachers should employ some didactic strategies that encourage students to perform stretching exercises outside the PE setting (Viciana et al., 2014). Future research interventions should examine the effectiveness of these kinds of interventions for maintaining the flexibility levels during the long holidays such as Christmas or summer. Additionally, future research studies should also examine in depth the effects of different maintenance stretching programs in order to retain the flexibility gains obtained previously.

Conclusion

In conclusion, the results of the present study first suggest that a short-term, four-minute development program improves students' hamstring extensibility, also increasing the proportion of schoolchildren with normal hamstring extensibility. Secondly, another important finding is the fact that it is possible to maintain students' hamstring extensibility only with one-minute sessions twice a week in a PE setting. Since only 2% of the PE time in a common situation of 50-minute sessions twice a week was used, apart from maintaining the flexibility levels previously obtained, these programs would not interfere in the normal teaching of other curricular PE contents. Finally, the results of this study show that schoolchildren with low hamstring extensibility improved greater than those with normal hamstring extensibility. This knowledge could help and guide teachers to design programs that allow a feasible and effective development and maintenance of students' flexibility in the PE setting.

Acknowledgements

Authors thank Aliisa Hatten-Viciana for the English revision. The first author is supported by the Spanish Ministry of Education, Culture and Sport (AP2010-5905).

References

Ayala, F., Sainz de Baranda, P., de Ste Croix, M. and Santonja, F. (2012) Fiabilidad y validez de las pruebas sit-and-reach: Revisión sistemática. *Revista Andaluza de Medicina del Deporte* **5,** 57–66. (In Spanish).

Carregaro, R.L. and Coury, H.J.C. (2009) Does reduced hamstring flexibility affect trunk and pelvic movement strategies during manual handling? *International Journal of Industrial Ergonomics* **39,** 115–120.

Castro-Piñero, J., Girela-Rejón, M.J., González-Montesinos, J.L., Mora, J., Conde-Caveda, J., Sjöström, M. and Ruiz, J.R. (2013) Percentile values for flexibility tests in youths aged 6 to 17 years: Influence of weight status. *European Journal of Sport Science* **13,** 139–148.

Coledam, D.H.C., Arruda, G.A. and Ramos de Oliveira, A. (2012) Chronic effect of static stretching performed during warm-up on flexibility in children. *Revista Brasileira de Cineantropometria e Desempenho Humano* **14,** 296–304.

DePino, G.M., Webright, W.G. and Arnold, B.L. (2000) Duration of maintained hamstring flexibility after cessation of an acute static stretching protocol. *Journal of Athletic Training* **35,** 56–59.

Díaz-Soler, M.A., Vaquero-Cristóbal, R., Espejo-Antúnez, L. and López-Miñarro, P.A. (2015) Efecto de un protocolo de calentamiento en la distancia alcanzada en el test sit and-reach en alumnos adolescentes. *Nutrición Hospitalaria* **31,** 2618–2623. (In Spanish).

European Commission/ EACEA/ Eurydice (2013) *Physical education and sport at school in Europe Eurydice Report*. Publications Office of the European Union, Luxembourg.

Feldman, D.E., Shrier, I., Rossignol, M. and Abenhaim, L. (2001) Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology* **154,** 30–36.

Ferrer, V. (1998) *Repercusiones de la cortedad isquiosural sobre la pelvis y el raquis* lumbar. Doctoral thesis, University of Murcia, Murcia. 1–277. (In Spanish).

Gajdosik, R.L. (2001) Passive extensibility of skeletal muscle: Review of the literature with clinical implications. *Clinical Biomechanics* **16,** 87–101.

Haley, S.M. and Fragala-Pinkham, M.A. (2006) Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy* **86,** 735–743.

Hardman, K. (2008) Physical education in schools: A global perspective. *Kinesiology* **40,** 5–28.

Hedges, L.V. (2007) Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics* **32,** 341–370.

Hoeger, W.W., Hopkins, D.R., Button, S. and Palmer, T.A. (1990) Comparing the sit and reach with the modified sit and reach in measuring flexibility in adolescents. *Pediatric Exercise Science* **2,** 156–162.

Jones, M.A., Stratton, G., Reilly, T. and Unnithan, V. (2005) Biological risk indicators for recurrent non-specific low back pain in adolescents. *British Journal of Sports Medicine* **39,** 137–140.

Kamandulis, S., Emeljanovas, A. and Skurvydas, A. (2013) Stretching exercise volume for flexibility enhancement in secondary school children. *The Journal of Sports Medicine and Physical Fitness* **53,** 687–692.

Kujala, U.M., Taimela, S., Salminen, J.J. and Oksanen, A. (1994) Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonathletes. *Scandinavian Journal of Medicine and Science in Sports* **4,** 200–205.

López-Miñarro, P.A., Muyor, J.M. and Alacid, F. (2012) Influence of hamstring extensibility on sagittal spinal curvatures and pelvic tilt in high-trained young kayakers. *European Journal of Sport Science* **12,** 469–474.

McHugh, M.P., Kremenic, I.J., Fox, M.B. and Gleim, G.W. (1998) The role of mechanical and neural restraints to joint range of motion during passive stretch. *Medicine and Science in Sports and Exercise* **30,** 928–932.

Mayorga-Vega, D., Merino-Marban, R. and Garcia-Romero, J.C. (2015) Validity of sit-and-reach with plantar flexion test in children aged 10-12 years. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte* **59,** 577–591.

Mayorga-Vega, D., Merino-Marban, R., Garrido, F.J. and Vician, J. (2014a) Comparison between warm-up and cool-down stretching programs on hamstring extensibility gains in primary schoolchildren. *Physical Activity Review* **2,** 16–24.

Mayorga-Vega, D., Merino-Marban, R., Vera-Estrada, F. and Viciana, J. (2014b) Effect of a short-term physical education-based flexibility program on hamstring and lumbar extensibility and its posterior reduction in primary schoolchildren. *Kinesiology* **46,** 227–233.

Mayorga-Vega, D., Merino-Marban, R. and Viciana, J. (2014c) Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine* **13,** 1–14.

Mayorga-Vega, D. and Viciana, J. (2015) Las clases de educación física solo mejoran la capacidad cardiorespiratoria de los alumnos con menor condición física: Un estudio de intervención controlado. *Nutrición Hospitalaria* **32,** 330–335. (In Spanish).

Mayorga-Vega, D., Viciana, J. and Cocca, A. (2013) Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics* **37,** 153–160.

Meredith, M.D. and Welk, G.J. (2010) *Fitnessgram® and Activitygram®. Test administration manual*. 4th edition. Human Kinetics, United States.

Merino-Marban, R., Mayorga-Vega, D., Fernandez-Rodriguez, E., Vera Estrada, F. and Viciana, J. (2015) Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *European Physical Education Review* **21,** 83–92.

Mikkelsson, L., Nupponen, H., Kaprio, J., Kautiainen, H., Mikkelsson, M. and Kujala, U.M. (2006) Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine* **40,** 107–113.

Ministerio de Educación y Ciencia (2006) Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación Primaria. *Boletín Oficinal del Estado* **293,** 43075–43080. (In Spanish).

National Association for Sport and Physical Education (2005) *Moving into the future: National standards for physical education*. 2nd edition. McGraw-Hill Higher Education, Reston.

Ortega, F.B., Ruiz, J.R., Castillo, M.J. and Sjöström, M. (2008) Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity* **32,** 1–11.

O'Sullivan, K., Murray, E. and Sainsbury, D. (2009) The effect of warm-up, static stretching and dynamic stretching on hamstring flexibility in previously injured subjects. *BMC Musculoskeletal Disorders* **10,** 37.

Rancour, J., Holmes, C.F. and Cipriani, D.J. (2009) The effects of intermittent stretching following a 4-week static stretching protocol: A randomized trial. *Journal of Strength and Conditioning Research* **23,** 2217–2222.

Rodríguez, P.L., Santonja, F.M., López-Miñarro, P.A., Sáinz de Baranda, P. and Yuste, J.L. (2008) Effect of physical education stretching programme on sit-and-reach score in schoolchildren. *Science and Sports* **23,** 170–175.

Ryan, E.D., Everett, K.L., Smith, D.B., Pollner, C., Thompson, B.J., Sobolewski, E.J. and Fiddler, R.E. (2014) Acute effects of different volumes of dynamic stretching on vertical jump performance, flexibility and muscular endurance. *Clinical Physiology and Functional Imaging* **34,** 485–492.

Sainz de Baranda, P. (2009) Flexibility training in physical education: Intervention program. *Cultura, Ciencia y Deporte* **4,** 33–38.

Sainz de Baranda, P., Rodríguez, P.L., Santonja, F.M., López, P.A., Andújar, P., Ferrer, V. and Pastor, A. (2006) Effects of hamstring stretching exercises on the toe-touch test in elementary schoolchildren. *Journal of Human Movement Studies* **51,** 277–289.

Sánchez Rivas, E., Mayorga-Vega, D., Fernández Rodríguez, E. and Merino-Marban, R. (2014) Efecto de un programa de estiramiento de la musculatura isquiosural en las clases de educación física en educación primaria. *Journal of Sport and Health Research* **6,** 159–168.

Santonja Medina, F.M., Sainz de Baranda Andújar, P., Rodríguez García, P.L., López Miñarro, P.A. and Canteras Jordana, M. (2007) Effects of frequency of static stretching on straight-leg raise in elementary school children. *Journal of Sports Medicine and Physical Fitness* **47,** 304–308.

Shrout, P.E. and Fleiss, J.L. (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* **86,** 420–428.

Siedentop, D. and Tanehill, D. (2000) *Developing teaching skills in physical education*. Mayfield Publishing Company, Mountain View.

Sjölie, A.N. (2004) Low back pain adolescents are associated with poor hip mobility and high body mass index. *Scandinavian Journal of Medicine and Science in Sports* **14,** 168–175.

Spernoga, S.G., Uh, T.L., Arnold, B.L. and Gansneder, B.M. (2001) Duration of maintained hamstring flexibility after a one-time, modified hold-relax stretching protocol. *Journal of Athletic Training* **36,** 44–48.

Thacker, S.B., Gilchrist, J., Stroup, D.F. and Kimsey, C.D. (2004) The impact of stretching on sports injury risk: A systematic review of the literature. *Medicine and Science in Sports and Exercise* **36,** 371–378.

Thomas, J.R., Nelson, J.K. and Silverman, S.J. (2015) *Research methods in physical activity*. 7th edition. Human Kinetics, Champaign.

Viciana, J. and Mayorga-Vega, D. (2015) Innovative teaching units applied to physical education. Changing the curriculum management for authentic outcomes. *Kinesiology*, in press.

Viciana, J., Mayorga-Vega, D. and Merino-Marban, R. (2014) Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In: *Handbook of physical education research. Role of school programs, children's attitudes and health implications*. Ed: Todaro, R. New York: Nova Science Publisher. 237–252.

Weppler, C.H. and Magnusson, S.P. (2010) Increasing muscle extensibility: A matter of increasing length or modifying sensation? *Physical Therapy* **90,** 438–449.

Willy, R.W., Kyle, B.A., Moore, S.A. and Chleboun, G.S. (2001) Effect of cessation and resumption of static hamstring muscle stretching on joint range of motion. *Journal of Orthopaedic and Sports Physical Therapy* **31,** 138–144.

Legends

Figure 1. Flow chart corresponding to the participants included in the present study.

Figure 2. The six stretching exercises performed during the physical education-based stretching program were: (a) standing with feet together; (b) sitting with feet together; (c) standing with feet shoulder-width apart; (d) sitting with feet shoulder-width apart; (e) standing with only one leg extended, and (f) sitting with only one leg extended.

Table 1. General characteristics (mean ± standard deviation/ frequency) of the participants and differences between the experimental and control groups.

Table 2. Effect of the physical education-based stretching program on hamstring extensibility (classic sit-and-reach scores, cm).

Table 3. Effect of the physical education-based stretching intervention program on the proportion of children with normal hamstring extensibility.

Figure 3. Effect of the physical education-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline (classic sit-and-reach scores, cm).

Key points

- A short-term stretching development program improves students' hamstring extensibility, also increasing the proportion of schoolchildren with normal hamstring extensibility.
- A physical education-based stretching maintenance program of only one-minute sessions twice a week is effective in maintaining hamstring extensibility among schoolchildren.
- A four-minute maintenance program shows similar effects that the one-minute maintenance program on hamstring extensibility among schoolchildren.
- After a stretching development program, schoolchildren with low hamstring extensibility improve more than those with normal hamstring extensibility.

Authors Biography

Daniel Mayorga-Vega

Degree

MSc

Affiliation

Department of Physical Education and Sport, University of Granada, Spain.

Research interests

Measurement and evaluation, health-related physical fitness, health-enhancing physical activity, physical education-based interventions, motivation toward physical activity.

Address

Alfacar street, 18011, Granada, Spain.

Phone number: 0034-958246641

Fax number: 0034-958244369

E-mail: dmayorgavega@gmail.com

Rafael Merino-Marban

Degree

PhD

Affiliation

Facultad de Ciencias de la Educación, Universidad de Málaga (UMA), Spain.

Research interests

Health-related physical fitness, physical education-based interventions, kinesio tape.

Address

Cervantes Av., 2. Campus of Teatinos, 29071 Malaga, Spain.

Phone number: 0034-952132464

Fax number: 0034-952134102

E-mail: rmerino@uma.es

Jorge Manzano-Lagunas

Degree

BSc

Affiliation

CEIP Virgen de Belén, Junta de Andalucía, Spain.

Research interests

Health-related physical fitness, physical education-based interventions.

Address

Virgen de Belén Ave, nº 7 and 10, 29004, Málaga, Spain.

Phone number: 0034-951298301

Fax number: 0034-951298302

E-mail: maestrojorgemanzano@gmail.com

Humberto Blanco

Degree

PhD

Affiliation

Faculty of Physical Education and Sport Sciences, Autonomous University of Chihuahua, Mexico.

Research interests

Educational technology and research in teaching, development of measuring instruments in education, health and social sciences.

Address

Universitaria campus, 2-1585, 31009, Chihuahua, Mexico.

Phone number: 614-1589900

Fax number: 614-1589900

E-mail: hblanco@uach.mx

Jesús Viciana

Degree

PhD

Affiliation

Department of Physical Education and Sport, University of Granada, Spain.

Research interests

Measurement and evaluation, health-related physical fitness, health-enhancing physical activity, physical education-based interventions, motivation toward physical activity.

Address

Alfacar street, 18011, Granada, Spain.

Phone number: 0034-958246641

Fax number: 0034-958244369

E-mail: jviciana@ugr.es

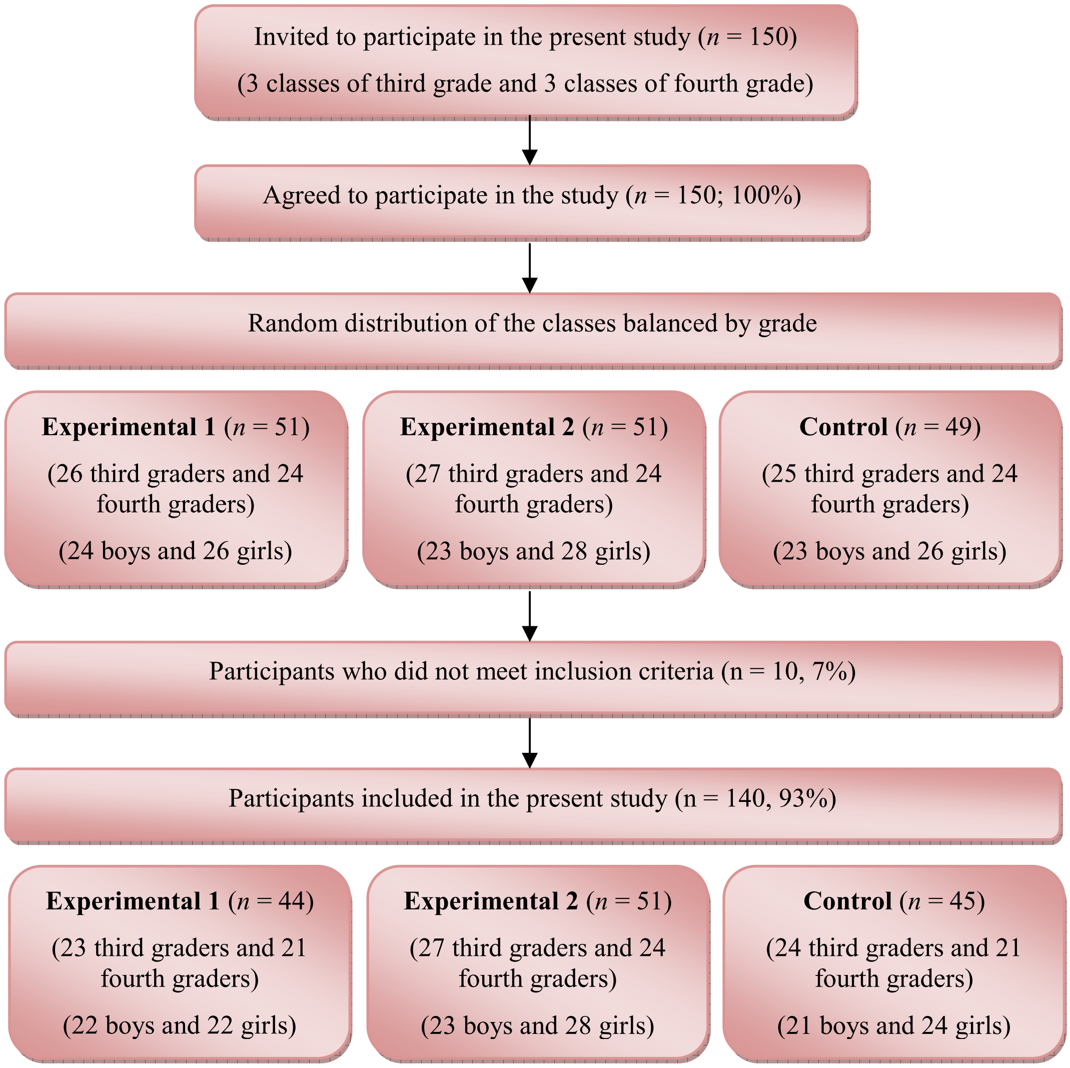


Figure 1. Flow chart corresponding to the participants included in the present study.

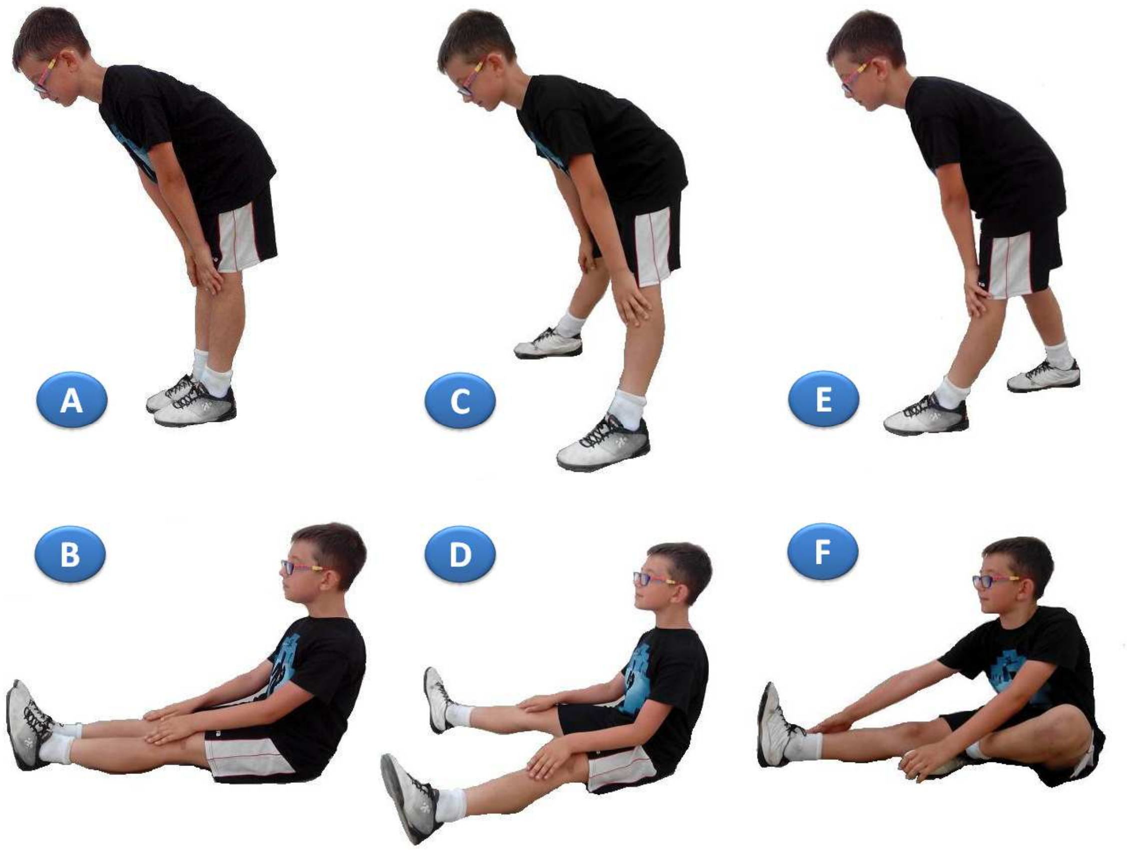


Figure 2. The six stretching exercises performed during the physical education-based stretching program were: (a) standing with feet together; (b) sitting with feet together; (c) standing with feet shoulder-width apart; (d) sitting with feet shoulder-width apart; (e) standing with only one leg extended, and (f) sitting with only one leg extended.

Table 1. General characteristics (mean ± standard deviation/ frequency) of the participants and differences between the experimental and control groups.

	Experimental 1 (n = 44)	Experimental 2 (n = 51)	Control (n = 45)	p^a
Age (years)	8.5 ± 0.8	8.4 ± 0.8	8.4 ± 0.6	-
Gender (boys/ girls)	22/ 22	23/ 28	21/ 24	0.890
Body mass (kg)	33.8 ± 7.5	33.7 ± 7.7	36.6 ± 11.5	0.229
Body height (cm)	133.2 ± 6.4	132.4 ± 6.1	132.6 ± 8.0	0.843
Body mass index (kg/ m^2)	19.0 ± 3.4	19.1 ± 3.4	20.4 ± 4.4	0.124
Extracurricular sport (yes/ no)[b]	33/ 11	28/ 23	25/ 20	0.082
Pre-intervention score (cm)	16.8 ± 5.4	16.8 ± 5.5	15.3 ± 5.2	0.320
Hamstring extensibility (low/ normal)	15/ 29	15/ 36	18/ 27	0.551

Note. [a] Significance level from the chi squared test for the ratio of gender, extracurricular sport, and pre-intervention hamstring extensibility categories, and from the one-way analysis of variance for body mass, body height, body mass index, and pre-intervention scores of the sit-and-reach test; [b] Children that regularly participated (yes) or not (no) at least twice per week in extracurricular sport activities.

Table 2. Effect of the physical education-based stretching program on hamstring extensibility (classic sit-and-reach scores, cm).

Group	Pre-inter. (1) ($M \pm SD$)	Post-dev. (2) ($M \pm SD$)	Post-mant. (3) ($M \pm SD$)	p[a]	Effect size[b]		
					1-2	2-3	1-3
EG1 (n = 44) ##	16.8 ± 5.4	19.5 ± 6.0***	20.1 ± 5.6***	< 0.001	0.49	0.03	0.52
EG2 (n = 51) #	16.8 ± 5.5	19.1 ± 5.1***	19.1 ± 5.0***		0.42	- 0.10	0.33
CG (n = 45)	15.3 ± 5.2	15.4 ± 4.9	15.8 ± 5.7		0.07	0.13	0.20

Note. M, mean; SD, standard deviation; EG1, experimental group 1; EG2, experimental group 2; CG, control group; [a] Significance level from the two-way analysis of variance with the *post hoc* analyses with Bonferroni adjustment. Between-groups pairwise comparisons: Differences statistically significant from CG (# p < 0.05, ## p < 0.01). Within-groups pairwise comparisons: Change statistically significant from pre-intervention to post-development and from post-development to post-maintenance (*** p < 0.001); [b] Hedges' g effect size. Rows from top to bottom: EG1-CG, EG2-CG, and EG1-EG2.

Table 3. Effect of the physical education-based stretching intervention program on the proportion of children with normal hamstring extensibility.

		Experimental group 1			Experimental group 2			Control group		
		Post-development			Post-development			Post-development		
		Low	Normal	Total	Low	Normal	Total	Low	Normal	Total
Pre-intervention	Low	7	8	15	4	11	15	14	4	18
		(15.9)	(18.2)	(34.1)	(7.8)	(21.6)	(29.4)	(31.1)	(8.9)	(40.0)
	Normal	1	28	29	2	34	36	6	21	27
		(2.3)	(63.6)	(65.9)	(3.9)	(66.7)	(70.6)	(13.3)	(46.7)	(60.0)
	Total	8	36	44	6	45	51	20	25	45
		(18.2)	(81.8)	(100.0)	(11.8)	(88.2)	(100.0)	(44.4)	(55.6)	(100.0)
	p^a			0.039			0.022			0.754
		Post-maintenance			Post-maintenance			Post-maintenance		
		Low	Normal	Total	Low	Normal	Total	Low	Normal	Total
Post -development	Low	4	4	8	2	4	6	17	3	20
		(9.1)	(9.1)	(18.2)	(3.9)	(7.8)	(11.8)	(37.8)	(6.7)	(44.4)
	Normal	2	34	36	7	38	45	2	23	25
		(4.5)	(77.3)	(81.8)	(13.7)	(74.5)	(88.2)	(4.4)	(51.1)	(55.6)
	Total	6	38	44	9	42	51	19	26	45
		(13.6)	(86.4)	(100.0)	(17.6)	(82.4)	(100.0)	(42.2)	(57.8)	(100.0)
	p^a			0.687			0.549			1.000

Note. Data are reported as frequency (% total). [a] Significance level from the McNemar's test.

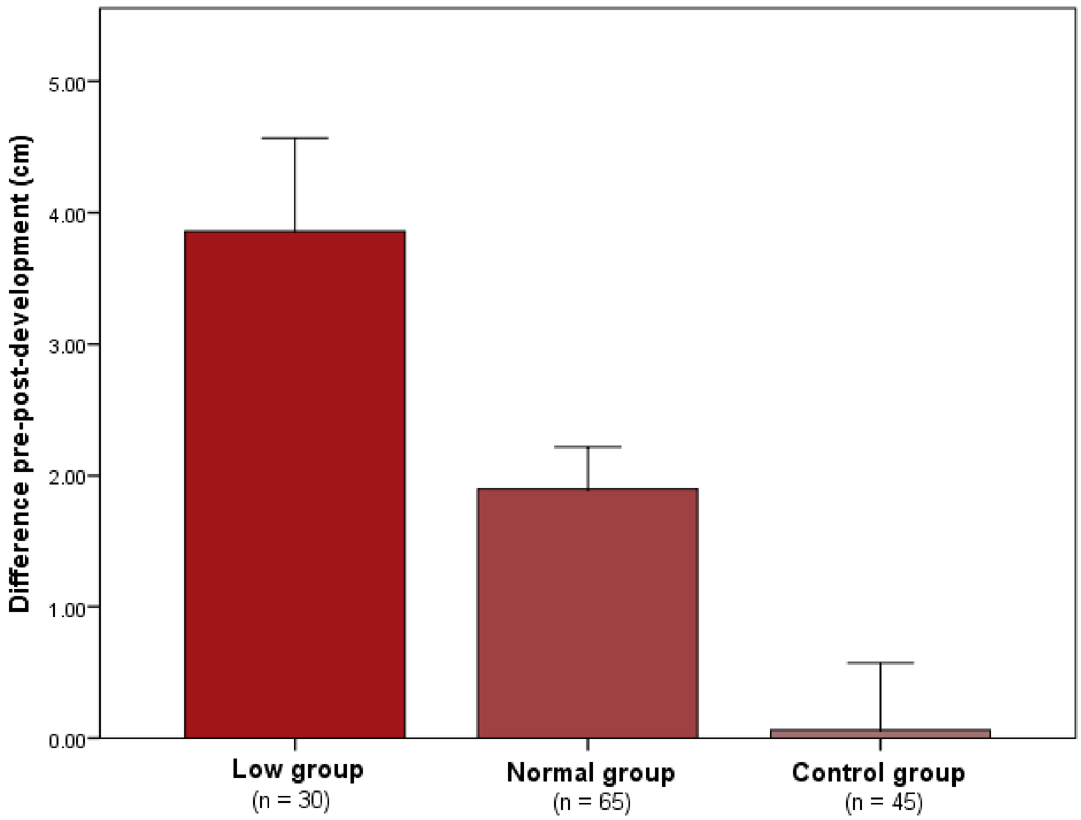


Figure 3. Effect of the physical education-based stretching intervention program on hamstring extensibility according to the children´s flexibility baseline (classic sit-and-reach scores, cm). The values represent the mean and the error bars the standard error.

5. EFFECTS OF PHYSICAL EDUCATION-BASED PHYSICAL FITNESS PROGRAMS ON PHYSICAL SELF-CONCEPT IN SCHOOLCHILDREN (PAPERS XV AND XVII)

Please see the paper XV in pages 329-346.

EFFECTS OF A PHYSICAL EDUCATION-BASED PROGRAM ON HEALTH-RELATED PHYSICAL FITNESS AND ITS MAINTENANCE IN HIGH SCHOOL STUDENTS: A CLUSTER-RANDOMIZED CONTROLLED TRIAL

Mayorga-Vega, D., Montoro-Escaño, J., Merino-Marban, R., & Viciana, J.

European Physical Education Review

In press

Paper XVII

EFFECT OF A PHYSICAL FITNESS PROGRAM ON PHYSICAL SELF-CONCEPT AND PHYSICAL FITNESS ELEMENTS IN PRIMARY SCHOOL STUDENTS

Mayorga-Vega, D., Viciana, J., Cocca, A., & De Rueda Villén, B.

Perceptual and Motor Skills

2012, *115*(3), 984-996.

Perceptual & Motor Skills: Exercise & Sport
2012, 115, 3, 984-996. © Perceptual & Motor Skills 2012

EFFECT OF A PHYSICAL FITNESS PROGRAM ON PHYSICAL SELF-CONCEPT AND PHYSICAL FITNESS ELEMENTS IN PRIMARY SCHOOL STUDENTS[1, 2]

D. MAYORGA-VEGA, J. VICIANA, A. COCCA, AND B. DE RUEDA VILLÉN

University of Granada, Spain

Summary.—The purpose of this study was to assess the short-term effect of a physical fitness program on physical self-concept and physical fitness elements among primary school students. Spanish boys and girls ($N = 75$; M age = 11.1 yr., $SD = 0.4$) were divided into an experimental group and a control group. During physical education classes, the experimental group performed an 8-week program including two circuits of 8 exercises done for 15 to 35 sec. each with 45 to 25 sec. of rest between them. Physical self-concept (Physical Self-Description Questionnaire) and physical fitness (EUROFIT battery tests) were measured at the beginning and at the end of the physical fitness program. The results showed that the improvements in physical fitness were not accompanied by major changes in physical self-concept, even though the physical fitness program seemed to maintain the Experimental group's previous physical appearance, strength, and self-esteem perceptions, all of which statistically significantly decreased in the control group after the intervention.

In recent decades, health has been regarded not merely as the freedom from disease or injury, but also a state of complete physical, mental, and social well-being (World Health Organization, 1946). Developed countries, out of concern for psychosocial pathologies associated with current aesthetic models, have promoted the health of schoolchildren from the physic/psychosocial dual reality (Ministry of Education, 2006). For this reason, schools, through the area of physical education (PE), have taken steps to make students aware of physical activity for their well-being and demonstrate a responsible attitude towards oneself and other persons, and recognize the healthy effects of physical activity (Spanish Ministry of Education, 2006).

Self-concept, defined as an individual's perception of him- or herself (Harter, 1990), is developed through experience and interpretation of the environment (Shavelson, Hubner, & Stanton, 1976). Self-concept gains great importance during the last stage of childhood, since it is at this stage when children experience major changes regarding their skills and their ability to assess their skills (Harter, 1999). Therefore, modifications of self-concept are more feasible during this period, because children at this age are more likely to use comparisons to evaluate their competence (Horn & Weiss, 1991) and thus to develop their self-concept.

---

[1]Address correspondence to J. Viciana, Ph.D., Department of Physical Education, University of Granada, Ctra. Alfacar s/n, 18011 Granada, Spain or e-mail (jviciana@ugr.es).
[2]The first author is supported by a grant from the Spanish Ministry of Education (AP2010-5905).

DOI 10.2466/06.10.25.PMS.115.6. 984-996 ISSN 0031-5125

One of the most important elements of overall self-concept in children is physical self-concept (Klesges, Haddock, Stein, Klesges, Eck, & Hanson, 1992), which has been defined as the perception individuals have of their physical skills and physical appearance (Stein, 1996). Physical self-concept is particularly relevant because of its connection with academic achievement (Gold, 1982; Ayora, García, & Rubio, 1997), but above all because of its effect on physical activity (Chan, Au, Chan, Kwan, Yiu, & Yeung, 2003; Planinsec & Fosnaric, 2005) and physical fitness (Chan, *et al.*, 2003; Carraro, Scarpa, & Ventura, 2010), with corresponding influences on health (Meredith & Dwyer, 1991), use of leisure time, and social relationships (Alfermann & Stoll, 2000).

At present, properly designed and competently monitored physical conditioning is widely regarded as a safe and effective method (American Academy of Pediatrics, 2008; Behm, Faigenbaum, Falk, & Klentrou, 2008; Faigenbaum, Kraemer, Blimkie, Jeffreys, Micheli, Nitka, *et al.*, 2009) for developing schoolchildren´s health-related physical fitness (Ortega, Ruiz, Castillo, & Sjöström, 2008). Varied studies have also found a positive relationship between health-related physical fitness and physical self-concept in children and adolescents (Ayora, *et al.*, 1997; Chan, *et al.*, 2003; Carraro, *et al.*, 2010).

However, changes in the physical self-concept of children as a result of a fitness program are limited and contradictory (Greene & Ignico, 1995; Faigenbaum, Zaichkowsky, Westcott, Long, LaRosa-Loud, Micheli, *et al.*, 1997; Sadres, Eliakim, Constantini, Lidor, & Falk, 2001). Most of the existing studies have focused on how sport affects physical self-concept (Weiss, McAulley, Ebbeck, & Wiese, 1990; Salokun, 1994). A smaller number of studies have taken physical fitness as an independent variable, especially in the PE setting. Because of the role played by school-based programs promoting the development of physical and psychosocial health-related markers, evaluation of the effectiveness of these kinds of programs is required. Currently, there is a lack of scientific information about the effect of short-term physical fitness programs on physical self-concept in a PE setting.

The purpose of present study was to evaluate the effects of an eight-week physical fitness program on the physical self-concept in primary education children in a PE setting. Two hypotheses were tested:

Hypothesis 1: A short-term physical fitness program will increase physical fitness in primary school students.

Hypothesis 2: A short-term physical fitness program will increase physical self-concept in primary school students.

Method

Participants

A sample of 75 Spanish primary schoolchildren volunteers, boys

($n = 41$) and girls ($n = 34$), participated in this study (*M* age = 11.1 yr., *SD* = 0.4; *M* body mass = 42.4 kg, *SD* = 9.95; *M* height = 146.5 cm, *SD* = 6.8; *M* Body Mass Index (BMI) = 19.7 kg/m^2, *SD* = 3.8). All students enrolled in Grade 6 at one school were invited to take part in the research. In addition, the following exclusion criteria were applied: (a) children with a chronic pediatric disease, (b) children with an orthopedic restriction, (c) children older than 12 years, and (d) children with previous experience in physical fitness programs.

All invited students agreed to participate and met the selection criteria, thus, all of them were accepted into the study. For practical reasons and because the research focused on reality and school context, a cluster randomized controlled trial was used (Sadres, *et al.*, 2001; Faigenbaum, Farrell, Fabiano, Radler, Naclerio, Ratamess, *et al.*, 2011). Natural school classes were assigned randomly to form the study groups (two classes for each group): (a) an experimental group ($n = 38$, 20 boys, 18 girls; *M* age = 11.1 yr., *SD* = 0.4; *M* body mass = 41.3 kg, *SD* = 9.3; *M* height = 144.6 cm, *SD* = 7.0; *M* BMI = 19.6 kg/m^2, *SD* = 3.5), who performed the fitness program for PE, and (b) a control group ($n = 37$, 21 boys, 16 girls; *M* age = 11.1 yr., *SD* = 0.4; *M* body mass = 45.2 kg, *SD* = 11.2; *M* height = 148.9 cm, *SD* = 5.7; *M* BMI = 20.2 kg/m^2, *SD* = 4.4) who attend standard PE classes. The two groups were balanced for sex ratio (47% and 43% girls, respectively).

The participants were urged to maintain their normal daily activities and allowed to participate in physical sport activities but not to carry out physical fitness training outside the supervised setting. Twenty-seven children in the Experimental (77%) and 27 children in the Control (73%) groups regularly participated (at least twice per week) in organized sport programs. The study protocol was also approved by The Ethical Committee of the University of Granada.

Measures

In order to check the effects of the fitness program, all students were evaluated one week before and one week after the intervention (pre-intervention and post-intervention, respectively). The physical self-concept evaluation was conducted using the Spanish version (Tomás, 1998) of the Physical Self-Description Questionnaire (PSDQ; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). The participants completed the questionnaires in their usual classroom with two members of the research team. Participants were asked to compare themselves on a scale from one to six with other children of the same age in terms of physical activity, motor ability, anthropometric characteristics, as well as other physical attributes. Researchers gave a printed copy of the questionnaires to each student, previously offering a general explanation on how it should be completed.

Before the children began to fill it in, the researchers resolved any questions. They were also present to answer questions that could arise during its completion.

Physical Self-Description Questionnaire (PSDQ).—This questionnaire consists of 70 items that measure nine specific components of physical self-concept (health, coordination, body fat, physical activity, sports competence, physical appearance, strength, flexibility, and endurance) and two global components (global physical self-concept and self-esteem). Participants compared themselves with other children of the same age. Each physical self-concept component was made up of six items, except for health and self-esteem, which had eight items. The response format is based on a 6-point Likert-type scale (1 = False, 2 = Mostly false, 3 = More false than true, 4 = More true than false, 5 = Mostly true, 6 = True; higher scores indicating higher physical self-concept). The scales had both positively and negatively worded questions. All negatively worded items (21 in total) are reverse scored and summed with other item ratings in the same scale. The Spanish scales have good internal consistency (Cronbach's alpha) ranging from .79 to .93 (Tomás, 1998).

Physical fitness.—Evaluation of the physical condition was done by means of the EUROFIT test battery (Council of Europe Committee for the Development of Sport, 1988), validated and standardized by the Council of Europe. The muscular strength and cardiovascular endurance EUROFIT tests were selected as being the best markers of health-related physical fitness among children (Ortega, *et al.*, 2008). Blind evaluations of children were carried out by two researchers, following the standard protocol for each test. Each researcher assessed physical fitness with the same tests using identical equipment. Prior to the evaluation, all participants performed a warm-up consisting of 5 min. OF running from low to moderate intensity. The order and a brief description of the test protocol are set out as follows.

The standing long jump was used to evaluate lower-limb explosive strength. The participants stood behind a starting line and were instructed to push off vigorously and jump as far as possible. The participants had to land with their feet together and to stay upright. The distance was measured from the take-off line to the point where the back of the heel nearest to the take-off line landed on the floor. A further attempt was allowed if the participants fell backward or touched the floor with another part of the body. The best score (cm) of two attempts was recorded.

Number of sit-ups in 30 sec. was used to assess abdominal endurance strength. The children lay in the supine position on a mat with their knees bent at 90° and their feet flat on the floor, held down by a researcher. The hands were placed at the back of the head, fingers interlaced. On hearing "Go!" the students' elbows had to contact the knees and return to the starting position as many times as possible in 30 sec. The total number of repetitions completed in 30 sec. was recorded.

The bent-arm hang was used to evaluate upper-limb endurance strength. The participants had to maintain a bent arm position while hanging from a bar with hands in a forward grip and at shoulder width. The participants' chin had to be above the bar and they had to hold this position as long as possible without resting their chin on the bar. The test ended when the participants' eyes sink below the bar. Total time (sec.) was recorded.

The 20-m shuttle run was the assessment of cardiovascular endurance. All students ran between two parallel lines 20 m apart, the rhythm marked by a recorded beep. A researcher ran alongside the children to help them keep the pace. The starting speed was 8.5 km/hr.; it increased 0.5 km/hr. every minute. The test ended when the child stopped running due to fatigue or failed to reach the line before the next signal for two consecutive times. The time of the last completed lap (sec.) was recorded.

Procedure

The Control participants had PE classes twice a week on non-consecutive days given by their usual teacher, according to Spanish Ministry of Education (2006) guidelines. The sole special feature of these classes during the research was that the students were not allowed to work on the PE subject "Physical activity and health." During these classes the Control students performed traditional games and an introduction to basketball and volleyball. The Experimental group carried out a fitness program during their PE classes, led by a researcher. The class was set up as a circuit, because it allowed for longer motor engagement time (Lozano, 2005). In addition, as several studies observed that feedback positively influences self-concept (Lozano, Cocca, Salinas, Miranda, & Viciana, 2007), each student received task-oriented and affective positive feedback at least once per session during the implementation of the program.

The intervention program was designed following the general recommendations for children (American Academy of Pediatrics, 2008; Behm, *et al.*, 2008; Faigenbaum, *et al.*, 2009). The students in the Experimental group performed program sessions twice per week on non-consecutive days for 8 weeks, led by a researcher. As there were school holidays on two of the PE dates, a total of 14 sessions were completed in the program. Each session was 50 min. long and consisted of 5 min. of warm-up racing games, the 40-min. fitness program, and a cool-down period with two 15- to 30-sec. sets of static stretching, mainly on hamstrings and lumbar muscles (Table 1).

The fitness program included two circuits with eight stages. Each

TABLE 1
INTERVENTION SESSION

Exercise	Intensive Progression (Levels 1/ 2/ 3)[a]	Material
Warm-up (5 min.)		
Racing games		
Main work-out (40 min.)		
Strength stations		
a. Throwing from chest	1kg/1.5kg/ 2kg	Medicine ball
b. Rowing	Low/medium/high resistance	Elastic band
c. Going up-down	Body weight/+1kg/+2kg	Swedish bench
d. Triceps extension	Low/ medium/high resistance	Elastic band
e. Biceps curl	Low/ medium/high resistance	Elastic band
f. Skipping rope	Micropause/with/without rebound	Rope
g. Crunches	Arms stretched forward/chest/backward	Mat
h. Bridging	Body weight/+1kg/+2kg	Mat, medicine ball
Aerobic station		
i. Racing games		
Cool-down (5 min.)		
Static stretching		

Note.—[a]All participants began at the first level of difficulty. When students could perform more than one repetition per second they were allowed to advance to next level (Weltman, *et al.*, 1986).

stage consisted of a 15- to 35-sec. exercise, with a rest time of 25 to 45 sec. between stages. The work and the rest times varied over the intervention following the principle of training progression. As suggested by Weltman, Janney, Rians, Strand, Berg, Tippitt, *et al.* (1986), during working time students had to complete as many repetitions as they could without intervention from the researcher. In addition, as indicated by previous studies, the last repetition of each series should show momentary muscle fatigue (Faigenbaum, Westcott, Micheli, Outerbridge, Long, LaRosa-Loud, *et al.*, 1996; Faigenbaum, Westcott, LaRosa-Loud, & Long, 1999; Faigenbaum, Milliken, LaRosa-Loud, Burak, Doherty, & Westcott, 2002; Faigenbaum, Milliken, Moulton, & Westcott, 2005). To this end, children were offered three difficulty levels for each stage, so that the exercise intensity was best fitted to each student's capacity. All participants began at the first level of difficulty. When a student could perform more than one repetition per second they were allowed to proceed to the next level (Weltman, *et al.*, 1986). With the aim of developing cardiovascular endurance, at the end of each circuit all students simultaneously executed an additional stage consisting of a 5-min. endurance racing game.

Statistical Analysis

Descriptive statistics were calculated on age, body mass, height, Body Mass Index (BMI), and the results of the different tests. As the bent-arm hang variable did not follow a normal distribution, the data were log-transformed (Bland & Altman, 1996). A 2 Group (Control, Experimental) × 2 Time (Pre-intervention, Post-intervention) multivariate analysis of covariance (MANCOVA), with pre-intervention values of the studied variables as covariates, was used to test the effect of the intervention program on the self-perception and EUROFIT results. Significant multivariate interactions were followed up with univariate analysis of covariance (ANCOVA). Subsequently, a Bonferroni adjustment was used to compare pairs of means. Effect size was estimated using the eta-squared (η^2). Internal consistency reliability was estimated using the intraclass correlation coefficient (ICC) from a two-way analysis of covariance (ANOVA; Shrout & Fleiss, 1979). Furthermore, 95% confidence intervals (*CI*) were calculated. All statistical analyses were performed using the SPSS Version 15.0 for Windows (SPSS® Inc., Chicago, IL). The statistical significance level was set at $p < .05$.

Results

Six of the initial 75 participants were excluded because they failed to meet the attendance required by the intervention program (at least 12 of the 14 scheduled meetings) or did not attend the post-intervention evaluation. Average attendance of the Experimental group participants was 94%. Table 2 shows the mean values and standard deviations of physical self-concept and physical fitness measures, as well as the results of MANCOVA and follow-up univariate ANCOVA.

Physical Self-concept

The MANCOVA results indicated overall significant interaction effect on PSDQ scales (Wilks' $\lambda = 0.67$, $p = .05$). Follow-up ANCOVA revealed significant differences in interaction effects on perception of health, physical appearance, strength, and self-esteem.[3] Then, the within-groups factor with Bonferroni adjustment showed that in the Experimental group this did not produce statistically significant differences in self-perception scores for health, physical appearance, strength, or self-esteem. In the Control group, the within-groups factor with Bonferroni adjustment showed a statistically significant decrease in self-perception scores on physical appearance, strength, and self-esteem. However, the Control group had higher scores on health perception at the end than at the beginning of the intervention period. For the other physical self-concept components, the

[3]The study determined whether the girls' scores decreased over the intervention in the control group and whether children who were engaged in sports activities were less affected by the fitness program. We did not find statistically significant differences in any of these comparisons (pre-post-intervention). Significant differences were found in the comparison between sports-engaged children and those who were not engaged in any sports activity at the baseline, with sports-engaged children scoring the highest values on physical self-concept.

TABLE 2

EFFECTS OF THE INTERVENTION PROGRAM ON PHYSICAL SELF-CONCEPT AND PHYSICAL FITNESS VALUES

Group × Pre/Post-intervention	Pre-intervention				Post-intervention				F	p	η^2
	Experimental ($n = 35$)		Control ($n = 34$)		Experimental ($n = 35$)		Control ($n = 34$)				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*			
PSDQ[a]									2.04	.05	0.33
Health	5.18	0.71	4.69	0.77	5.03	0.71	5.04†	0.81	7.33	.01	0.17
Coordination	4.57	0.99	4.70	0.86	4.42	1.09	4.55	0.98	0.25	.62	0.00
Physical activity	4.49	1.13	4.16	1.45	4.24	1.25	4.22	1.38	0.56	.46	0.01
Body fat	5.03	1.13	4.79	1.38	5.07	1.18	4.65	1.51	2.39	.13	0.04
Sports competence	4.42	1.11	4.49	0.91	4.38	1.26	4.25*	1.21	1.63	.21	0.03
Physical appearance	4.45	0.98	4.37	1.21	4.56	1.04	3.96†	1.52	9.67	.003	0.15
Strength	4.12	1.25	4.27	1.14	4.15	1.46	3.94†	1.24	6.14	.02	0.10
Flexibility	4.03	1.37	4.10	1.13	3.72	1.29*	3.93	1.25	0.19	.67	0.00
Endurance	3.96	1.08	3.88	1.17	3.81	1.23	3.53†	1.46	2.59	.13	0.04
Global physical self-concept	5.31	0.89	5.21	0.93	5.16	1.04	4.89*	1.39	0.66	.42	0.01
Self-esteem	5.09	0.89	4.97	0.85	5.18	0.89	4.76*	1.13	4.76	.03	0.08
EUROFIT									3.56	.01	0.20
Bent arm hang[b] (sec.)	11.59	10.18	17.42	19.80	15.11‡	13.07	15.07	14.72	5.04	.03	0.08
Sit-ups in 30 sec. (number)	20.20	4.39	17.59	4.76	21.77†	3.82	17.36	6.10	9.73	.003	0.14
Standing long jump (cm)	126.66	24.98	123.82	21.92	127.29	25.32	127.42	25.73	0.10	.76	0.00
20-m shuttle run (sec.)	164.09	94.05	192.99	99.99	200.73*	99.54	185.88	98.53	4.05	.05	0.06

Note.—[a]In the PSDQ (Marsh, *et al*., 1994) the scores range from 1 to 6 on each item, the higher values corresponding to better perception of each variable; [b]For statistical analysis the data were log-transformed. *$p < .05$. †$p < .01$. ‡$p < .001$ for within-group changes.

ANCOVA results revealed no significant interaction effects. The internal consistency [ICC (95%*CI*)] for the overall PSDQ ranged from .76 (.66, .84) to .94 (.92, .96).

Physical Fitness

The MANCOVA results indicated an overall significant interaction effect on physical fitness elements (Wilks' $\lambda = 0.80$, $p = .01$). Follow-up ANCOVA revealed significant differences interaction effects in the bent-arm hang, sit-ups in 30 sec., and 20-m shuttle run (Table 2). Then, the Experimental group's within-groups factor with Bonferroni adjustment showed statistically significant changes in the bent-arm hang, sit-ups in 30 seconds, and 20-m shuttle run. On the other hand, the Control group's within-groups factor with Bonferroni adjustment did not show statistically significant changes in the bent-arm hang, sit-ups in 30 seconds, or 20-m shuttle run. For the standing long jump the follow-up ANCOVA did not reveal interaction effects. The test-retest reliabilities [ICC (95% CI)] of the physical fitness measures were: bent-arm hang .91 (.81, .95), sit-ups in 30 sec. .85 (.70, .93), standing long jump .93 (.85, .96), and 20-m shuttle run .90 (.80, .95).

Discussion

A physical fitness program in the PE context with Spanish students of primary education, organized as a circuit over eight weeks with a frequency of two weekly sessions, was effective in improving muscular endurance and cardiovascular endurance. Several studies have obtained significant results after implementing similar after-school fitness programs (8 weeks, twice a week) with school children of the same age (Faigenbaum, *et al.*, 1996, 1999, 2002, 2005). Conducted outside the school context, all these studies took a number of maximum repetitions as reference for their tests. However, the present study was conducted according to work-rest time, more appropriate in the school context (Weltman, *et al.*, 1986).

In addition, in the PE setting it is more advisable to program the tasks on the basis of working time for the teachers to monitor the whole group. This also enables the class to be organized in work circuits with multiple concurrent tasks that notably increase motor engagement time (Lozano, 2005), a matter of utmost importance in limited session programs. Along these lines, a study of Weltman, *et al.* (1986), which consisted of circuit resistance training (14 weeks, three times a week, three circuits of 30 seconds work-rest), obtained significant gains for both muscle strength and cardiovascular endurance.

In the present study the explosive strength of the legs did not show any change in the Experimental group. One possible reason for this result, in contrast to the gains for muscular endurance, may lie in the differences between the test and training methods. The present intervention was strength endurance training, that is, the ability to generate continuous strength for long periods, especially when loads are low (American Academy of Pediatrics, 2008). This is different from muscular explosive strength, which is the capacity to generate the maximum strength in the shortest time (American Academy of Pediatrics, 2008). Although both expressions refer to strength, they are measures with different physiological and metabolic characteristics (Behm, *et al.*, 2008; Faigenbaum, *et al.*, 2009). Other studies with similar designs also have failed to find significant changes in explosive strength of the legs as measured by the standing long jump (Faigenbaum, *et al.*, 1996, 2002, 2005).

Studies with adolescents and adults have found a positive influence of physical fitness programs on physical self-concept (Brazell-Roberts & Thomas, 1990; Tucker, 1983; Velez, Golem, & Arent, 2010). However, in the present study the PSDQ results in the Experimental group did not show statistically significant changes. This is in line with previous studies where the programs carried out in children did not produce any change in their physical self-concept (Faigenbaum, *et al.*, 1997; Sadres, *et al.*, 2001), or changed only in some dimensions (Greene & Ignico, 1995). In addition, the values in both groups were high from the beginning (between four and five, on a scale from one to six), which were at least maintained during the study, as they were in other investigations carried out in the PE context (Sadres, *et al.*, 2001). These high basal values are consistent with those obtained in longitudinal studies, which show an increase towards higher values of self-concept at the end of primary education (Cole, Maxwell, Martin, Peeke, Seroczynski, Tram, *et al.*, 2001); such scores seem difficult to increase as a result of a short-term program. In contrast, during the intervening period, the Control group participants showed a significant decrease in the perception of physical appearance, strength, and self-esteem.

The Control students had higher scores on health perception at the end than at the beginning of the intervention period; on this set of items, health is regarded as not getting sick often and getting well quickly when you are sick (Marsh, *et al.*, 1994). Hence, in this questionnaire children are asked to evaluate their health as the absence of illness, and not a state of complete physical well-being, which is related with objective health-related physical fitness (Marsh, 1996; Guérin, Marsh, & Famose, 2004; Carraro, *et al.*, 2010). In addition, the Experimental group had statistically significantly higher scores on health perception at baseline, so that it does not seem that they can be increased by a short-term physical fitness program.

In conclusion, this study suggests that a short-term physical fitness program (8 weeks) organized as a circuit within the context of the PE period (just two weekly sessions) and applied to primary education children, can improve cardiovascular and muscular endurance. However, the changes this intervention causes in physical fitness are not accompanied by major changes in physical self-concept, even though this intervention seems to maintain previous levels for some of the dimensions that produce a decrease in the control group. A limitation of the present study was that the sample was not large enough for gender and age subgroups to be formed and their results compared. Likewise, not combining the fitness program with some sport practice, or the program's short duration may have limited the effects of this study on physical self-concept. Regarding the feedback effect, all teachers in PE need to motivate students during the tasks (students in both groups in this study were encouraged), both to achieve the aim of the subject as well as from the ethical point of view. To control the feedback effect, a second control group without any kind of feedback could be used. For these reasons, future research where these suggestions can be incorporated is required.

REFERENCES

Alfermann, D., & Stoll, O. (2000) Effects of physical exercise on self-concept and well-being. *International Journal of Sport Psychology*, 31, 47-65.

American Academy of Pediatrics. (2008) Strength training by children and adolescents. *Pediatrics*, 121, 835-840.

Ayora, D., García, A., & Rubio, S. (1997) Factores del autoconcepto relacionados con el rendimiento de los adolescentes en educación física [The self-concept factors related to performance of adolescents in physical education]. *Revista de Psicología del Deporte*, 6, 59-73.

Behm, D. G., Faigenbaum, A. D., Falk, B., & Klentrou, P. (2008) Canadian Society for Exercise Physiology position paper: resistance training in children and adolescents. *Applied Physiology, Nutrition, and Metabolism*, 33, 547-561.

Bland, J. M., & Altman, D. (1996) Transforming data. *British Medical Journal*, 312, 770.

Brazell-Roberts, J. V., & Thomas, L. E. (1990) Pumping iron and self-esteem: starting a weight-training program increased the reported self-esteem of college-aged women. *Swimming Technique*, 26, 21-23.

Carraro, A., Scarpa, S., & Ventura, L. (2010) Relationships between physical self-concept and physical fitness in Italian adolescents. *Perceptual and Motor Skills*, 110, 522-530.

Chan, E. W. C., Au, E. Y. M., Chan, B. H. T., Kwan, M. K. M., Yiu, P. Y. P., & Yeung, E. W. (2003) Relations among physical activity, physical fitness, and self-perceived fitness in Hong Kong adolescents. *Perceptual and Motor Skills*, 96, 787-797.

Cole, D. A., Maxwell, S. E., Martin, J. M., Peeke, L. G., Seroczynski, A. D., Tram, J. M., Hoffman, K. B., Ruiz, M. D., Jacquez, F., & Maschman, T. (2001) The development of multiple domains of child and adolescent self-concept: a cohort sequential longitudinal design. *Child Development*, 72, 1723-1746.

Council of Europe Committee for the Development of Sport. (1988) *EUROFIT: handbook for the EUROFIT Tests of Physical Fitness*. Rome: Edigraf Editorial Grafica.

Faigenbaum, A. D., Farrell, A., Fabiano, A., Radler, T., Naclerio, F., Ratamess, N. A., Kang, J., & Myer, G. D. (2011) Effects of integrative neuromuscular training on fitness performance in children. *Pediatric Exercise Science*, 23, 573-584.

Faigenbaum, A. D., Kraemer, W. J., Blimkie, C. J. R., Jeffreys, I., Micheli, L. J., Nitka, M., & Rowland, T. W. (2009) Youth resistance training: updated position statement paper from the National Strength and Conditioning Association. *Journal of Strength and Conditioning Research*, 23, 60-79.

Faigenbaum, A. D., Milliken, L. A., LaRosa-Loud, R., Burak, B. T., Doherty, C. L., & Westcott, W. L. (2002) Comparison of 1 and 2 days per week of strength training in children. *Research Quarterly for Exercise and Sport*, 73, 416-424.

Faigenbaum, A. D., Milliken, L. A., Moulton, L., & Westcott, W. L. (2005) Early muscular fitness adaptations in children in response to two different resistance training regimens. *Pediatric Exercise Science*, 17, 237-248.

Faigenbaum, A. D., Westcott, W. L., LaRosa-Loud, R., & Long, C. (1999) The effects of different resistance training protocols on muscular strength and endurance development in children. *Pediatrics*, 104, 1-7.

Faigenbaum, A. D., Westcott, W. L., Micheli, L. J., Outerbridge, A. R., Long, C. J., LaRosa-Loud, R., & Zaichkowsky, L. D. (1996) The effects of strength training and detraininig on children. *Journal of Strength and Conditional Research*, 10, 109-114.

Faigenbaum, A., Zaichkowsky, L. D., Westcott, W. L., Long, C. J., LaRosa-Loud, R., Micheli, L. J., & Outerbridge, A. R. (1997) Psychological effects of strength training on children. *Journal of Sport Behavior*, 20, 164-175.

Gold, M. (1982) Scholastic experiences, self-esteem, and delinquent behavior: a theory for alternative schools. In M. Rosenberg & H. B. Kaplan (Ed.), *Social psychology of the self-concept*. Arlington Heights, IL: Harlan Davidson. Pp. 440-451.

Greene, J. C., & Ignico, A. A. (1995) The effect of a ten-week physical fitness program on fitness profiles, self-concept, and body-esteem in children. *Journal of the International Council for Health, Physical Education, Recreation, Sport, and Dance*, 31, 42-47.

Guérin, F., Marsh, H. W., & Famose, J. P. (2004) Generalizability of the PSDQ and its relationship to physical fitness: the European French connection. *Journal of Sport & Exercise Psychology*, 26, 19-38.

Harter, S. (1990) Issues in the assessment of the self-concept of children and adolescents. In A. M. La Greca (Ed.), *Through the eyes of the child: obtaining self-reports from children and adolescents*. Boston: Allyn & Bacon. Pp. 292-325.

Harter, S. (1999) *The construction of self: a developmental perspective*. New York: Guilford.

Horn, T. S., & Weiss, M. R. (1991) A developmental analysis of children's self-ability judgements in the physical domain. *Pediatric Exercise Science*, 3, 310-326.

Klesges, G., Haddock, C. K., Stein, R. J., Klesges, L. M., Eck, L. H., & Hanson, C. L. (1992) Relationship between psychosocial functioning and body fat in preschool children: a longitudinal investigation. *Journal of Consulting & Clinical Psychology*, 60, 793-796.

Lozano, L. (2005) *The influence of three systems of organization on motor engagement time in the classes of physical education with high school students*. Granada, Spain: Univer. of Granada.

Lozano, L., Cocca, A., Salinas, F., Miranda, M. T., & Viciana J. (2007) The physical self-concept of swimmers from other sports. In R. Arellano, J. A. Sánchez, F. Navarro, E. Morales, & G. López (Ed.), *Swimming science I*. Granada, Spain: Publications service of the Univer. of Granada. Pp. 203-208.

Marsh, H. W. (1996) Construct validity of physical self-description questionnaire responses: relations to external criteria. *Journal of Sport & Exercise Psychology*, 18, 111-131.

MARSH, H. W., RICHARDS, G. E., JOHNSON, S., ROCHE, L., & TREMAYNE, P. (1994) Physical self description questionnaire: psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology*, 16, 270-305.

MEREDITH, C. M., & DWYER, J. T. (1991) Nutrition and exercise: effects on adolescent health. *Annual Reviews Public Health*, 12, 309-333.

ORTEGA, F. B., RUIZ, J. R., CASTILLO, M. J., & SJÖSTRÖSM, M. (2008) Physical fitness in childhood and adolescence: a powerful marker of health. *International Journal of Obesity*, 32, 1-11.

PLANINSEC, J., & FOSNARIC, S. (2005) Relationship of perceived physical self-concept and physical activity level and sex among young children. *Perceptual and Motor Skills*, 100, 349-353.

SADRES, E., ELIAKIM, A., CONSTANTINI, N., LIDOR, R., & FALK, B. (2001) The effect of long-term resistance training on anthropometric measures, muscle strength, and self concept in pre-pubertal boys. *Pediatric Exercise Science*, 13, 357-372.

SALOKUN, S. O. (1994) Positive change in self-concept as a function of improved performance in sports. *Perceptual and Motor Skills*, 78, 752-754.

SHAVELSON, R. J., HUBNER, J. J., & STANTON, G. C. (1976) Self-concept: validation of construct interpretations. *Review of Educational Research*, 46, 407-441.

SHROUT, P. E., & FLEISS, J. L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

SPANISH MINISTRY OF EDUCATION. (2006) *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la educación primaria* [*Royal Decree 1513/2006, 7 December, which establishes the minimum teaching requirements of primary education*]. Madrid: Boletín Oficial del Estado.

STEIN, R. J. (1996) Physical self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept*. New York: Wiley. Pp. 374-394.

TOMÁS, I. (1998) [*Psychometric equivalence of a translation of the PSDQ (Physical Self-Description Questionnaire) into Spanish*]. Valencia, Spain: Univer. of Valencia. [in Spanish]

TUCKER, L. A. (1983) Effect of weight training on self-concept: a profile of those influenced most. *Research Quarterly for Exercise and Sport*, 54, 389-397.

VELEZ, A., GOLEM, D. L., & ARENT, S. M. (2010) The impact of a 12-week resistance training program on strength, body composition, and self-concept of hispanic adolescents. *Journal of Strength and Conditional Research*, 24, 1065-1073.

WEISS, M. R., MCAULEY, E., EBBECK, V., & WIESE, D. M. (1990) Self-esteem and causal attributions for children's physical and social competence in sport. *Journal of Sport and Exercise Psychology*, 12, 21-36.

WELTMAN, A., JANNEY, C., RIANS, C. B., STRAND, K., BERG, B., TIPPITT, S., WISE, J., CAHILL, B. R., & KATCH, F. I. (1986) The effects of hydraulic resistance strength training in pre-pubertal males. *Medicine & Science in Sports & Exercise*, 18, 629-638.

WORLD HEALTH ORGANIZATION. (1946) *Constitution of the World Health Organization*. New York: International Health Conference.

Accepted November 8, 2012.

6. INFLUENCE OF MOTIVATION TOWARD PHYSICAL EDUCATION ON THE EFFECTIVENESS OF INTERVENTION PROGRAMS FOR IMPROVING STUDENTS´ HEALTH-RELATED PHYSICAL FITNESS (PAPER XVIII)

Paper XVIII

¿LA MOTIVACIÓN AUTODETERMINADA HACIA LA EDUCACIÓN FÍSICA INFLUYE EN LA MEJORA DE LA CONDICIÓN FÍSICA SALUDABLE?

[DOES THE SELF-DETERMINATION MOTIVATION TOWARD PHYSICAL EDUCATION INFLUENCE THE PHYSICAL FITNESS IMPROVING?]

Mayorga-Vega, D., Montoro-Escaño, J., & Viciana, J.

Revista de Psicología del Deporte

Submitted

¿La motivación autodeterminada hacia la educación física influye en la mejora de la condición física?

Does the self-determination motivation toward physical education influence the physical fitness improving?

Resumen

El objetivo del presente estudio fue examinar la influencia de la motivación autodeterminada hacia la educación física sobre la eficacia de un programa de intervención durante las clases de educación física para la mejora de la condición física en estudiantes de educación secundaria. Una muestra de 97 estudiantes de educación secundaria obligatoria fue asignada aleatoriamente (por clases naturales) al grupo experimental y control. Durante las clases de educación física, los estudiantes del grupo experimental realizaron un programa de acondicionamiento físico. Los resultados del análisis de covarianza de un factor mostraron que los estudiantes experimentales con moderada y alta motivación autodeterminada hacia la educación física incrementaron estadísticamente su condición física con respecto a los estudiantes controles (p < 0,001). En cambio, no se encontraron diferencias estadísticamente significativas entre los estudiantes experimentales con baja motivación hacia la educación física y los controles (p > 0,05). Un programa de acondicionamiento físico durante las clases de educación física solo parece incrementar la condición física de aquellos estudiantes con moderada-alta motivación autodeterminada hacia la educación física. Por ello, con el objetivo de incrementar la condición física saludable de sus alumnos, los profesores de educación física deberían promover la motivación hacia la educación física.

Palabras clave: Teoría de la autodeterminación, programa de acondicionamiento físico, capacidad cardiorespiratoria, intervención educativa, adolescentes.

Abstract

The purpose of the present study was to examine the influence of self-determined motivation toward physical education on the effectiveness of an intervention program during physical education classes for improving physical fitness in high school students. A sample of 97 high school students was randomly assigned (by natural groups) to the experimental and control groups. During physical education classes, students of the experimental group performed a physical fitness program. The one-way analysis of covariance results showed that experimental students with moderate and high self-determined motivation toward physical education statistically improved their physical fitness comparing with control students ($p < 0.001$). However, no statistically significant differences between the experimental students with low motivation toward physical education and control students were found ($p > 0.05$). A physical fitness program during physical education classes only seems to improve the physical fitness of students with moderate-high self-determined motivation toward physical education. Therefore, in order to increase the health-related physical fitness of their students, physical education teachers should promote motivation toward physical education.

Key words: Self-determination theory, physical fitness program, cardiorespiratory fitness, educational intervention, adolescents.

¿La motivación autodeterminada hacia la educación física influye en la mejora de la condición física?

Hoy en día la condición física es considerada uno de los indicadores más potentes de salud y calidad de vida entre los niños y adolescentes (Ortega, Ruiz, Castillo, y Sjöström, 2008; Ruiz et al., 2009). Por ejemplo, en la actualidad hay evidencia científica que revela que, entre los jóvenes, tener un mayor nivel de condición física se asocia con mejores resultados de salud, tales como tener un menor riesgo presente y futuro de obesidad, un menor riesgo de padecer enfermedades cardiovasculares durante la vida adulta y un esqueleto más saludable (Ortega et al., 2008; Ruiz et al., 2009). Además entre los jóvenes, mejorar los niveles de condición física también parece tener efectos positivos sobre marcadores psicológicos de salud tales como depresión, ansiedad, estado de ánimo y autoestima (Ortega et al., 2008).

Por tanto, las políticas de promoción de la salud también deberían estar diseñadas para identificar a los niños con bajos niveles de condición física, así como fomentar su mejora mediante la promoción de niveles adecuados de actividad física (Ortega et al., 2008). En este sentido las escuelas, a través de la asignatura de educación física (EF) principalmente, podrían jugar un papel muy importante en esta cuestión. En muchos países, la EF es una asignatura obligatoria para todos los estudiantes (Hardman, 2008). Además, normalmente los profesores de EF deben llevar a cabo programas con el fin de mejorar los niveles de condición física de sus estudiantes (Ministerio de Educación y Ciencia, 2007; National Association for Sport and Physical Education, 2004). Para ello se recomienda que durante las clases de EF los estudiantes acumulen al menos el 50% de su tiempo en actividades de intensidad moderada-vigorosa (United States Department of Health and Human Services, 2010).

Sin embargo, el potencial de la asignatura de EF está marcadamente restringida por su limitada asignación de tiempo curricular (European Commission/ EACEA/ Eurydice, 2013), especialmente durante la etapa de educación secundaria obligatoria donde la frecuencia y duración de las clases se reducen a menudo en beneficio de otras áreas curriculares. Por otro lado, el éxito de la asignatura también se verá limitado si los estudiantes no están motivados para participar activamente en sus clases de EF (Ntoumanis, 2001). La motivación es un rasgo psicológico que despierta al individuo para actuar hacia una meta deseada y provoca, controla y sostiene ciertos comportamientos dirigidos a un objetivo; por tanto, la motivación es el propósito o causa psicológica de cualquier acción (Ryan, de Williams, Patrick, y Deci, 2009). En este sentido, numerosos estudios han encontraron que el criterio mínimo de actividad física durante las clases de EF anteriormente mencionado rara vez se cumple (Fairclough y Stratton, 2005, 2006).

La teoría de la autodeterminación (Deci y Ryan, 1985) constituye un marco conceptual especialmente útil para comprender las consecuencias de la motivación hacia la EF (Ntoumanis, 2001). Dicha teoría adopta una perspectiva multidimensional de la motivación, distinguiendo entre las razones de “por qué” los individuos se animan a actuar. La teoría de la autodeterminación sugiere que la regulación del comportamiento hacia una actividad se podría ordenar en un *continuum* de acuerdo con el grado en que la motivación es autodeterminada: intrínsecamente motivado (autodeterminado), extrínsecamente motivado (controlado), o desmotivado (sin intención) (Deci y Ryan, 2000). La motivación intrínseca está relacionada con el placer y la satisfacción inherente proporcionada por la propia actividad. Por el contrario, la motivación extrínseca se caracteriza porque la actividad se realiza para obtener algunos resultados independientes como, por ejemplo, recompensas tangibles, evitar castigos, o alcanzar reconocimientos o aprobaciones. Por último, la desmotivación se caracteriza por la no intención del individuo a actuar debido a diferentes razones tales como la falta de percepción de competencia o la ausencia de conexión entre la acción y los resultados deseados (Ryan et al., 2009).

Numerosos estudios anteriores han encontrado cómo entre los adolescentes la motivación autodeterminada hacia la EF está positivamente relacionada con sus niveles de actividad física (e.g., Kalaja, Jaakkola, y Liukkonen, 2010; Ullrich-Francés y Cox, 2009; Yli-Piipari, Watt, Jaakkola, Liukkonen, y Nurmi, 2009). De modo similar, los resultados de varios estudios previos parecen indicar que los estudiantes con mayor motivación autodeterminada hacia la EF también son físicamente más activos durante sus clases de EF que aquellos con menor motivación (Lonsdale, Sabiston, Raedeke, Ha, y Sum, 2009; Mayorga-Vega y Viciana, 2014). Sin embargo, lamentablemente no se han encontrado estudios previos que analicen el rol de la motivación hacia la EF de los jóvenes sobre el efecto de un programa para la mejora de la condición física. Consecuentemente, el objetivo del presente estudio fue examinar la influencia de la motivación autodeterminada hacia la EF sobre la eficacia de un programa de intervención durante las clases de EF para la mejora de la condición física en estudiantes de educación secundaria.

Método

Participantes

Una muestra de 97 estudiantes, 62 niños y 35 niñas, de seis clases de 2º curso de un centro de educación secundaria obligatoria participaron en el presente estudio. Por razones prácticas y debido a la naturaleza del presente estudio, se empleó un diseño *cluster-randomized controlled trial* (Mayorga-Vega, Viciana, y Cocca, 2013; Merino-Marban, Mayorga-Vega, Fernandez-Rodriguez, Vera Estrada, y Viciana, 2015). Las seis clases naturales se asignaron aleatoriamente al grupo control y el grupo experimental.

Todos los participantes estaban libres de padecer cualquier trastorno de salud que le impidieran realizar actividad física como tener enfermedades del corazón, asma no controlada o problemas osteoarticulares. El criterio de inclusión fue tener una asistencia a las clases de EF del programa de intervención igual o superior al 85%. Todos los estudiantes y sus tutores legales fueron plenamente informados acerca de las características del estudio y firmaron su consentimiento informado. El protocolo del estudio fue aprobado por el Comité Ética de la Universidad de [eliminado por cuestiones de anonimato].

Instrumentos

Motivación hacia la EF. La evaluación de la motivación hacia la EF se llevó a cabo en un aula regular antes del comienzo del programa de intervención, usando la versión española de la Escala del Locus Percibido de Causalidad en EF (Goudas, Biddle, y Fox, 1994). El cuestionario consta de un total de 20 ítems repartidos en cinco dimensiones que miden la motivación intrínseca, tres formas de regulación de la motivación extrínseca (identificada, introyectada y externa) y la desmotivación. Para cada dimensión, cuatro ítems se clasifican en una escala tipo Likert de 7 puntos con las etiquetas de 1 = "Totalmente en desacuerdo" hasta 7 = "Totalmente de acuerdo". El cuestionario está encabezado por la frase "Participo en esta clase de EF…", y los ítems representan las diferentes razones de dicho enunciado, reflejando los diferentes tipos de motivación.

Posteriormente, para la medida de la motivación autodeterminada, se calculó el índice de motivación autodeterminada mediante la ponderación de las puntuaciones de las diferentes dimensiones: (2 × motivación intrínseca + regulación identificada) - ((regulación introyectada + regulación externa) / 2 + 2 × desmotivación) (Vallerand y Rousseau, 2001). El uso de este índice se debe al hecho de que distintos tipos de motivación están teóricamente posicionados en un *continuum* de autodeterminación que va desde la motivación intrínseca a la desmotivación (Vallerand, Fortier, y Gaya, 1997). Por tanto, el índice de motivación autodeterminada integra las puntuaciones en las dimensiones de la motivación en una sola puntuación correspondiente a la posición de cada participante en dicho *continuum*. Un índice más positivo indica una mayor motivación autodeterminada hacia la EF y viceversa. La versión española de la Escala del Locus Percibido de Causalidad en EF tiene una propiedades psicométricas adecuadas (CFI = 0,89; RMSEA = 0,04; los coeficientes α de Cronbach oscilan de 0,7 a 0,8) (Moreno Murcia, González-Cutre, y Chillón Garzón, 2009).

Condición física. La evaluación de la condición física se llevó a cabo durante las clases de EF al comienzo y al final del programa de intervención (pre-intervención y post-intervención, respectivamente). Cada evaluación fue realizada por el mismo evaluador, instrumentos y condiciones. Las medidas fueron tomadas en una instalación deportiva cubierta con suelo antideslizante, en las mismas condiciones medioambientales, y a la misma hora y día de la semana para cada estudiante. Además, antes de la evaluación todos los participantes realizaron un calentamiento estandarizado de cinco minutos de carrera de baja a moderada intensidad.

La condición física de los alumnos se midió mediante el test Course Navette (Léger, Mercier, Gadoury, y Lambert, 1988). Los participantes corrieron entre dos líneas paralelas a 20 m de distancia al ritmo marcado por un sonido grabado. El test comenzaba a una velocidad de 8,5 km/ h e incrementaba 0,5 km/ h cada minuto. La prueba terminaba cuando los participantes dejaban de correr debido a la fatiga o cuando no lograban llegar a la línea antes de la siguiente señal en dos ocasiones consecutivas. Durante la realización de la prueba la frecuencia cardiaca de cada participante fue registrada mediante un pulsómetro (Polar® RS300X, Finlandia). Cada participante realizó la prueba una vez y se retuvo el tiempo total en segundos. Posteriormente, solo se utilizaron las marcas de aquellos participantes que alcanzaron el 90% o más de la frecuencia cardiaca estimada. El test Course Navette ha demostrado unos valores adecuados de validez entre los adolescentes (Léger et al., 1988).

Procedimiento

Los estudiantes del grupo experimental realizaron un programa de acondicionamiento físico durante las clases de EF dos veces por semana. Aunque el programa de intervención tuvo una duración de 22 semanas (primer y segundo trimestre), debido a diferentes motivos (vacaciones de navidad, días festivos puntuales y actividades académicas programadas), los estudiantes experimentales completaron un total de 30 sesiones. Cada sesión de intervención duró 50 minutos aproximadamente y consistió en un período de calentamiento de 5-10 minutos, 35-40 minutos de parte principal, y 5 minutos de vuelta a la calma. Durante la parte principal, los estudiantes experimentales realizaron sesiones de acondicionamiento físico comúnmente usadas en EF (e.g., juegos de carreras, entrenamiento en circuito, multisaltos, multi-lanzamientos, o entrenamiento integrado) seguidas por juegos o tareas de equipo. Durante las sesiones experimentales se puso especial énfasis en alcanzar una intensidad moderada a vigorosa.

Por su parte, durante el periodo de intervención, los estudiantes controles también participaron en sus dos sesiones semanales de EF, con una duración y estructura temporal similar. Sin embargo, los contenidos y metodologías seguidas durante la parte principal de las sesiones fueron diferentes a las sesiones experimentales. En cuanto a los contenidos, los estudiantes controles realizaron sesiones de deportes (voleibol, bádminton, baloncesto, hockey y acrosport). Por otro lado, se siguió una metodología tradicional de la EF poniendo un especial énfasis en el aprendizaje técnico-táctico en lugar de en la intensidad de la tarea.

Análisis estadístico

Se calcularon las medias y desviaciones estándar de la edad, masa corporal, talla e índice de masa corporal. Se utilizó el análisis de la varianza (ANOVA) de un factor, seguido por las comparaciones por pares con la corrección de Bonferroni, para estudiar las posibles diferencias en las características de los grupos. Posteriormente, el efecto del programa de intervención se estudió mediante un análisis de covarianza (ANCOVA) de un factor, incluyendo grupo como factor fijo, valores pre-intervención como covariable, y cambio pre-post intervención como variable dependiente. A continuación, se realizaron comparaciones por pares con la corrección de Bonferroni. Basado en estudios anteriores (Lonsdale et al., 2009), los estudiantes experimentales fueron previamente clasificados en función de su motivación autodeterminada hacia la EF: Baja (primer tertil), Moderada (segundo tertil) y Alta (tercer tertil). Por último, el tamaño del efecto *g* de Hedges se utilizó para estimar la magnitud del efecto del programa de intervención (Hedges, 2007). Todos los análisis estadísticos se realizaron mediante el paquete estadístico SPSS versión 20.0 para Windows (IBM® SPSS® Statistics 20). El nivel de significación estadística se estableció en $p < 0,05$.

Resultados

Los estudiantes experimentales obtuvieron una asistencia al programa de intervención media del 96-98%. La Tabla 1 muestra las características generales de los participantes del presente estudio. Los resultados del ANOVA de un factor no mostraron diferencias estadísticamente significativas entre los grupos estudiados ($p > 0{,}05$). La Figura 1 representa la motivación autodeterminada hacia la EF en función del grupo. El resultado del ANOVA de un factor mostró diferencias estadísticamente significativas entre los grupos estudiados [$F(3, 93) = 9{,}970$; $p < 0{,}001$; $\eta^2_p = 0{,}243$; P = 0,998]. Posteriormente, las comparaciones por pares con la corrección de Bonferroni mostraron que los estudiantes del grupo con alta motivación presentaban estadísticamente mayores valores que los estudiantes controles y con baja motivación ($p < 0{,}05$). Los estudiantes con moderada motivación presentaban estadísticamente mayores valores que los estudiantes con baja motivación ($p < 0{,}05$). Por otro lado, los estudiantes con baja motivación presentaban estadísticamente menores valores que los estudiantes controles ($p < 0{,}01$). Para el resto de comparaciones no se encontraron diferencias estadísticamente significativas ($p > 0{,}05$).

Insertar Tabla 1/ Insertar Figura 1

La Figura 2 muestra el efecto del programa de intervención sobre la condición física en función de la motivación autodeterminada hacia la EF de los estudiantes. Los resultados del ANCOVA de un factor sobre los valores obtenidos en la prueba de Course Navette mostraron un efecto estadísticamente significativo [$F(3, 92) = 10{,}379$; $p < 0{,}001$; $\eta^2_p = 0{,}253$; P = 0,998]. Posteriormente, las comparaciones por pares con la corrección de Bonferroni mostraron que los estudiantes con moderada y alta motivación hacia la EF incrementaron estadísticamente su condición física con respecto a los estudiantes controles ($p < 0{,}001$; $g = 0{,}46$). Los estudiantes con moderada y alta motivación hacia la EF también mostraron un incremento estadísticamente significativo con respecto a los estudiantes con baja motivación ($p < 0{,}05$; $g = 0{,}38$). En cambio, no se encontraron diferencias estadísticamente significativas entre los estudiantes con baja motivación hacia la EF y los estudiantes controles ($p > 0{,}05$; $g = 0{,}08$).

Insertar Figura 2

Discusión

El objetivo del presente estudio fue examinar la influencia de la motivación autodeterminada hacia la EF sobre la eficacia de un programa de intervención durante las clases de EF para la mejora de la condición física en estudiantes de educación secundaria. Hoy en día la asignatura de EF podría jugar un papel muy importante en cuestiones de salud pública, por ejemplo, mediante la mejora de los niveles de condición física saludable de los escolares (Ministerio de Educación y Ciencia, 2007; National Association for Sport and Physical Education, 2004). Sin embargo, en numerosos países el potencial de la asignatura de EF está notablemente restringida por diferentes limitaciones (Viciana, Mayorga-Vega, y Cocca, 2014; Viciana, Mayorga-Vega, y Merino-Marban, 2014) como, por ejemplo, su baja asignación de tiempo curricular (European Commission/ EACEA/ Eurydice, 2013).

A pesar de sus numerosas limitaciones, algunos estudios previos han encontrado que un programa de acondicionamiento físico durante las clases de EF dos veces por semana mejora la condición física de los escolares (Mayorga-Vega et al., 2013; Merino-Marban et al., 2015; Ramírez Lechuga et al., 2012). Sin embargo, el éxito de estos programas podrían verse seriamente limitados para aquellos estudiantes que no están motivados para participar activamente durante sus clases de EF (Ntoumanis, 2001). En este sentido, los resultados del presente estudio mostraron cómo solo los estudiantes con moderada y alta motivación autodeterminada hacia la EF incrementaron estadísticamente su condición física. En cambio, la intervención no incrementó la condición física de aquellos alumnos con baja motivación hacia la EF. En cuanto a la magnitud del efecto de la intervención, el tamaño del efecto del programa de acondicionamiento físico fue moderado para los estudiantes con moderada y alta motivación autodeterminada hacia la EF, mientras que para los de baja motivación la magnitud del efecto fue trivial. Estos hallazgos indican que el programa de intervención del presente estudio fue solo efectivo para aquellos alumnos con moderada y alta motivación autodeterminada hacia la EF.

Lamentablemente no se han encontrado estudios previos que analicen el rol de la motivación hacia la EF de los jóvenes sobre el efecto de un programa en la mejora de la condición física. No obstante, algunos estudios previos compararon los niveles objetivos de actividad física de los estudiantes durante las clases de EF en función de su motivación hacia la EF (Lonsdale et al., 2009; Mayorga-Vega y Viciana, 2014). De manera similar al presente estudio, dichas investigaciones encontraron que los estudiantes con mayor motivación autodeterminada hacia la EF también era físicamente más activos durante sus clases de EF que aquellos con menor motivación (Lonsdale et al., 2009; Mayorga-Vega y Viciana, 2014). Por otro lado, Chatzisarantis y Hagger (2008) estudiaron el efecto de un programa de intervención en EF basado en la teoría de la autodeterminación. Dichos autores encontraron que la intervención aumentó los valores de actividad física durante el tiempo libre reportado por los alumnos, así como la intención de ser físicamente más activos. Aunque las investigaciones anteriores se centraron en la actividad física en vez de la condición física, dado que los niveles de actividad física son un determinante principal para la mejora de la condición física, estos estudios apoyan los hallazgos del presente estudio.

En conclusión, de lo que conocemos, el presente estudio es el primero en comparar el efecto de un programa de intervención durante las clases de EF sobre la condición física en función del grado de motivación autodeterminada hacia la EF de sus estudiantes. Los resultados de este estudio sugieren que en un programa de acondicionamiento físico durante las clases de EF solo los estudiantes con moderada y alta motivación autodeterminada hacia la EF incrementan su condición física. En cambio, la intervención no incrementa la condición física de aquellos los alumnos con baja motivación autodeterminada hacia la EF. Por ello, con el objetivo de incrementar la condición física saludable de sus alumnos, los profesores de EF deberían promover la motivación autodeterminada hacia la EF. Estudios futuros deberían examinar el efecto de estimular la motivación autodeterminada hacia la EF de los estudiantes durante los programas de acondicionamiento físico en la mejora de la condición física.

Referencias

Chatzisarantis, N. L. D., y Hagger, M. S. (2009). Effects of an intervention based on self-determination theory on self-reported leisure-time physical activity participation. *Psychology and Health*, *24*, 29-48. doi: 10.1080/08870440701809533.

Deci, E. L., y Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., y Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227-268. doi: 10.1207/S15327965PLI1104_01

European Commission/ EACEA/ Eurydice (2013). *Physical education and sport at school in Europe Eurydice Report*. Luxembourg: Publications Office of the European Union.

Fairclough, S., y Stratton, G. (2005). Physical activity levels in middle and high school physical education: A review. *Pediatric Exercise Science*, *17*, 217-236.

Fairclough, S., y Stratton, G. (2006). A review of physical activity levels during elementary school physical education. *Journal of Teaching in Physical Education*, *25*, 239-257.

Goudas, M., Biddle, S. J. H., y Fox, K. (1994). Perceived locus of causality, goal orientations and perceived competence in school physical education classes. *British Journal of Educational Psychology, 64,* 453-463. doi: 10.1111/j.2044-8279.1994.tb01116.x

Hardman, K. (2008). Physical education in schools: A global perspective. *Kinesiology*, *40*, 5-28.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341-370. doi: 10.3102/1076998606298043

Kajala, S., Jaakkola, T., y Liukkonen, J. (2010). Fundamental movement skills and motivational factors influencing engagement in physical activity. *Perceptual & Motor Skills*, *111*, 115-128. doi: 10.2466/06.10.25.PMS.111.4.115-128

Léger, L. A., Mercier, D., Gadoury, C., y Lambert, J. (1988). The multistage 20 meter shuttle run test for aerobic fitness. *Journal of Sports Science*s, *6*, 93-101. doi: 10.1080/02640418808729800

Lonsdale, C., Sabiston, C. M., Raedeke, T. D., Ha, A. S. C., y Sum, R. K. W. (2009). Self-determined motivation and students´ physical activity during structured physical education lessons and free choice periods. *Preventive Medicine*, *48*, 69-73. doi: 10.1016/j.ypmed.2008.09.013

Mayorga-Vega, D., y Viciana, J. (2014). Adolescents´ physical activity in physical education, school recess, and extra-curricular sport by motivational profiles. *Perceptual and Motor Skills*, *118*, 663-679. doi: 10.2466/06.10.PMS.118k26w0

Mayorga-Vega, D., Viciana, J., y Cocca, A. (2013). Effects of a circuit training program on muscular and cardiovascular endurance and their maintenance in schoolchildren. *Journal of Human Kinetics*, *37*, 153-160. doi: 10.2478/hukin-2013-0036

Merino-Marban, R., Mayorga-Vega, D., Fernandez-Rodriguez, E., Vera Estrada, F., y Viciana, J. (2015). Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. *European Physical Education Review*, *21*, 83-92. doi: 10.1177/1356336X14550942

Ministerio de Educación y Ciencia (2007). Real Decreto 1613/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la educación secundaria obligatoria. *Boletín Oficial del Estado*, *5*, 677-773.

Moreno Murcia, J. A., González-Cutre, D., y Chillón Garzón, M. (2009). Preliminary validation in Spanish of a scale designed to measure motivation in physical education classes: The Perceived Locus of Causality (PLOC) Scale. *The Spanish Journal of Psychology*, *12*, 327-337. doi: 10.1017/S1138741600001724

National Association for Sport and Physical Education (2004). *Moving into the future: National Standards for Physical Education*. 2nd ed. Reston, VA: Author.

Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology*, *71*, 225-242. doi: 10.1348/000709901158497

Ortega, F. B., Ruiz, J. R., Castillo, M. J., y Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, *32*, 1-11. doi: 10.1038/sj.ijo.0803774

Ramírez Lechuga, J., Muros Molina, J. J., Morente Sánchez, J., Sánchez Muñoz, C., Femia Marzo, P., y Zabala Díaz, M. (2012). Effect of a 8-week aerobic training program during physical education lessons on aerobic fitness in adolescents. *Nutrición Hospitalaria*, *27*, 747-754. doi: 10.3305/nh.2012.27.3.5725

Ruiz, J. R., Castro-Piñero, J., Artero, E. G., Ortega, F. B., Sjöström, M., Suni, J., y Castillo, M. J. (2009). Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine*, *43*, 909-923. doi: 10.1136/bjsm.2008.056499

Ryan, R. M., Williams, G. C., Patrick, H., y Deci, E. L. (2009). Self-determination theory and physical activity: The dynamics of motivation in development and wellness. *Hellenic Journal of Psychology*, *6*, 107-124.

Ullrich-French, S., y Cox, A. (2009). Using cluster analysis to examine the combinations of motivation regulations of physical education students. *Journal of Sport and Exercise Psychology*, *31*, 358-379.

United States Department of Health and Human Services (2010). *Strategies to improve the quality of physical education*. Washington: Centers for Disease Control and Prevention.

Vallerand, R. J., Fortier, M. S., y Gaya, F. (1997). Self-determination and persistence in a real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology*, *72*, 1161-1176.

Vallerand, R. J., y Rousseau, F. L. (2001). Intrinsic and extrinsic motivation in sport and exercise: A review using the hierarchical model of intrinsic and extrinsic motivation. En R. N. Singer, H. A. Hausenblas, y C. M. Janelle (Eds.), *Handbook of sport psychology* (pp. 389-416). New York: Wiley.

Viciana, J., Mayorga-Vega, D., y Cocca, A. (2014) Successful learning model in physical education and its maintenance. The intermittent reinforcement applied to physical fitness. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, *9*, 155-171.

Viciana, J., Mayorga-Vega, D., y Merino-Marban, R. (2014). Physical education-based planning for developing and maintaining students' health-related physical fitness levels. En Todaro R (Ed.), *Handbook of physical education research. Role of school programs, children´s attitudes and health implications* (pp. 237-252). New York: Nova Science Publisher.

Yli-Piipari, S., Watt, A., Jaakkola, T., Liukkonen, J., y Nurmi, J. (2009). Relationship between physical education students' motivational profiles, enjoyment, state anxiety, and self-reported physical activity. *Journal of Sports Science & Medicine*, *8*, 327-336.

Agradecimientos

[Eliminado por cuestiones de anonimato].

Tabla 1

Características generales (medias ± desviaciones estándar) de los estudiantes

	Control (n = 47)	Baja[a] (n = 16)	Moderada[a] (n = 17)	Alta[a] (n = 17)
Edad (años)	12,5 ± 0,7	12,5 ± 0,6	12,4 ± 0,5	12,4 ± 0,6
Masa corporal (kg)	54,8 ± 17,1	49,7 ± 13,7	49,5 ± 7,7	54,9 ± 11,6
Talla (cm)	158,9 ± 7,6	155,1 ± 8,9	156,2 ± 6,7	161,1 ± 6,6
Índice de masa corporal (kg/m^2)	21,5 ± 5,6	20,4 ± 4,5	20,3 ± 2,9	21,1 ± 3,6

Nota. [a] Motivación autodeterminada hacia la educación física: Baja (primer tertil), Moderada (segundo tertil) y Alta (tercer tertil).

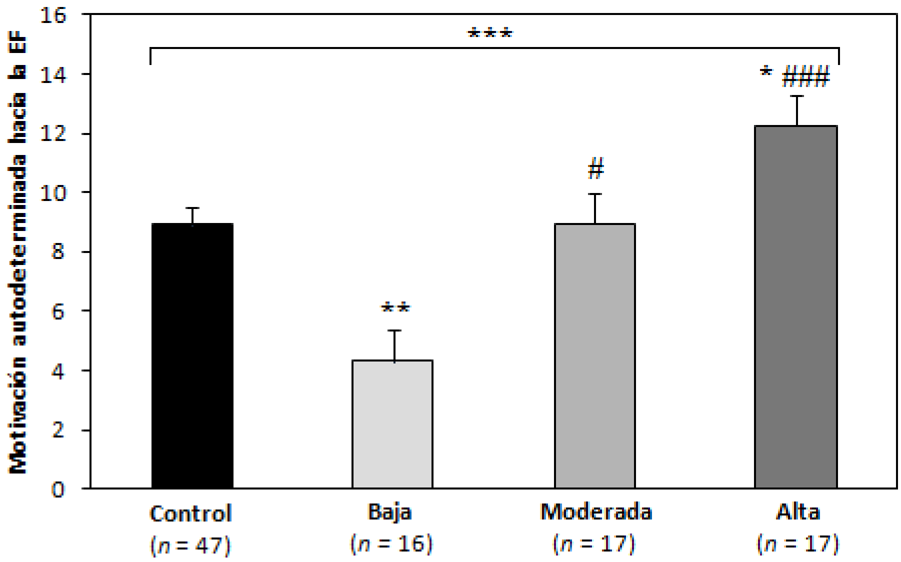


Figura 1. Motivación autodeterminada hacia la educación física en función del grupo. Los valores representan la media y las barras de error el error estándar. Nivel de significación del análisis de varianza de un factor (*** $p < 0{,}001$; sobre caja superior) seguido por las comparaciones por pares con la corrección de Bonferroni (sobre barras individuales): Diferencia estadísticamente significativa Control-Baja (** $p < 0{,}01$), Control-Alta (* $p < 0{,}05$), Baja-Moderada (# $p < 0{,}05$) y Baja-Alta (### $p < 0{,}001$).

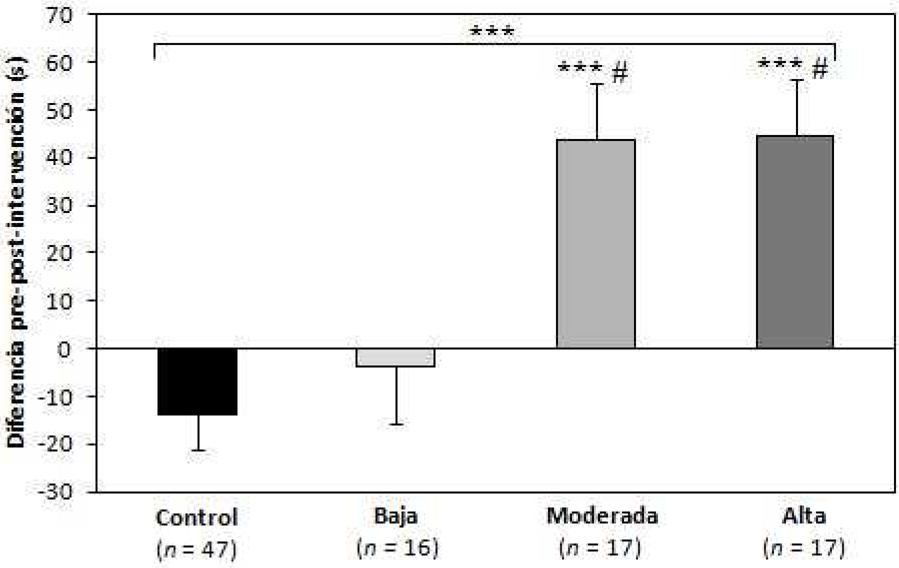


Figura 2. Efecto del programa de intervención sobre la condición física (s) en función de la motivación autodeterminada hacia la educación física de los estudiantes. Los valores representan la media ajustada y las barras de error el error estándar. Nivel de significación del análisis de covarianza de un factor (*** p < 0,001; sobre caja superior) seguido por las comparaciones por pares con la corrección de Bonferroni (sobre barras individuales): Cambio estadísticamente significativo Control-Moderada/ Control-Alta (*** p < 0,001) y Baja-Moderada/ Baja-Alta (# p < 0,05).

LIMITATIONS AND FUTURE RESEARCH STUDIES

[LIMITACIONES Y ESTUDIOS FUTUROS]

LIMITATIONS AND FUTURE RESEARCH STUDIES [LIMITACIONES Y ESTUDIOS FUTUROS]

The main limitations and future research studies of the present Doctoral Thesis are summarized below. For further information of any study, please check the corresponding section of each paper.

Paper I: As in any research study with a documental analysis of content, the definition of the representative subject categories of concepts that appear in the analyzed documents involves a subjective process (Viciana et al., 2007). Also noteworthy is that the absolute comparison of the frequencies of the covered topics in the both curriculum standards is highly influenced by the length of each document. However, the comparison by the percentage of subject categories could have solved slightly this mismatch between the two documents (Viciana Ramínez et al., 2007). Due to the apparition of a new national standard (Ministerio de Educación, Cultura y Deporte, 2014, 2015), future research studies should analysis and compare the content of the curriculum standards of the *Ley Orgánica para la Mejora de la Calidad Educativa* (LOMCE) and LOE in both primary and secondary education.

Paper II: Although a total of 211 pre-service PE teachers participated in the study, 31 planning had to be eliminated for showing important deficiencies that, following the *ad hoc* developed coding instrument, it did not allow an adequate analysis. Thus, the real percentage of PE planning with important deficiencies could be even higher than the reported in this study. Secondly, in spite the fact that the proposed index is a novel objective tool for examining the effectiveness of planning health-related physical fitness in PE, planning with key deficiencies with which students would not improve their health-related physical fitness levels could have a quite good score. Therefore, futures research studies should revise the proposed index. Finally, another limitation of the present study was related with the population studied. Due to practical reasons the planning of numerous in-service PE teachers was not accessible and those of pre-service teachers had to be analyzed instead. Although analyzing primary documents is very valuable, the main difficulty of such studies lies with the fact that the documents are private and, therefore, often causing suspicion to teachers to be studied (Thomas, Nelson, & Silverman, 2015). In order to analyze and identify potential deficiencies in the real planning of health-related physical fitness, future research studies should be conducted among in-service PE teachers. These findings would allow the government and educational institutions to know the real status of the PE planning, to approach this important national health issue in case of necessity.

Papers V-VIII: The main limitations of the present meta-analyses were related to the small number of criterion-related validity coefficients found. Estimating the population parameters based on small samples is simply less accurate than in a large-sized meta-analysis (Schmidt & Hunter, 2015). Additionally, because of a partial hierarchical breakdown (instead of full) had to be used, misleading results due to confounding and interaction effects might be produced (Schmidt & Hunter, 2015). Therefore, when a greater number of studies are accumulated, large sized meta-analyses with a full hierarchical analysis approach should be carried out.

Another potential limitation of these meta-analyses could be related to the statistical metric used. Some authors have argued that the correlation coefficient is a measure of *relationship* rather than *agreement* (Atkinson & Nevill, 1998). On the one hand, the performance score of the field tests (i.e., distance, time or speed) and the criterion measure (i.e., VO_2max or range of motion) are expressed in two different units and, therefore, an agreement statistical approach could not be performed. In order to solve this potential methodological limitation, another kind of validity such as the cross-validity or criterion-referenced validity could be followed instead (Baumgartner et al., 2015). Future research studies should examine the cross-validity and criterion-referenced validity of the analyzed physical fitness field tests. Furthermore, the validity of other type of cardiorespiratory fitness field tests such as the walk or step tests, flexibility field tests of other body regions and commonly used strength field tests should be also examined.

Another limitation of the present meta-analyses is related to the criterion measure used in the studies. For instance, although in the meta-analyses of cardiorespiratory fitness field tests only primary studies in which the criterion measure used the VO_2max relative to body mass during a laboratory-based incremental test to exhaustion were selected, researchers employed different equipment (various brand and characteristics), ergometers (i.e., treadmill and cycle ergometer) and protocols (e.g., in warm-up, initial load, increasing load, gas collection time, or number of gas collections). Furthermore, in primary studies researchers employed different criteria to determine VO_2max; a plateau in oxygen uptake, the respiratory exchange ratio, or the age-adjusted estimates of the maximal heart rate, alone or in combination were used; then, the quantitative cut-off values criteria were also diverse. Quantifying and accounting for these differences between studies remains a substantial methodological problem and a continuing source of debate (Cooper et al., 2009).

Finally, coding some study features was problematic due to different reasons. For instance, moderator analyses had missing data because authors mixed categories in their studies. Other study features simply could not be coded because the authors did not report them or they were ambiguous. Although authors were tried to contact, many of them did not reply and the particular study features had to be omitted. Because in the present meta-analysis the fitness level was classified based on the average scores, readers have to be aware that several individuals with low fitness could be classified as high fitness and vice versa. Lastly, although there could be other potentially moderating features such as physical activity levels or sport participation, coding for them was not possible because most studies did not report them.

Papers IX-XVIII: In the intervention studies there were some potential limitations that it should highlighted. Regarding the study design, the randomized control trial is often considered the "gold standard" to test the effectiveness of an intervention (Thomas et al., 2015). However, for practical reasons and the nature of the present studies, a cluster randomized controlled trial was used instead. Since the main objective of the present intervention studies was to examine the effect of the PE-based interventions in order to obtain direct practical applications in the setting, performing the programs in natural groups was a must. Therefore, although the cluster randomized controlled trial is a quasi-experimental design where the randomization is performed by natural groups instead of individuals such as in a true experimental design, the external validity is markedly greater (Thomas et al., 2015). Additionally, it must be mentioned that in schools commonly students are randomly assigned to the different classes. The fact that these intervention studies were always controlled by equivalent groups must be also highlighted. In this line, no study showed different statistically significant in general characteristics of the participants between experimental and control groups (except in the paper IX in which the body mass index was slightly different between groups in the sample of secondary school students). Statistical control was also used when it was necessary. Lastly, it should be highlighted that when classes from different grades were selected for a particular study, grades were randomly balanced to experimental and control groups.

Similarly, in papers IX, XVI (second objective) and XVIII, besides the manipulate variable by the researcher (i.e., the intervention program), and second condition could not be manipulated (*ex post fact* design). *Ex post facto* is a quasi-experimental design examining how an independent variable, present prior to the study, affects a dependent variable. That it is to say, *ex post facto* design does not include any form of manipulation before the fact occurs, as it is the case of a true experimental design. Therefore, although this design has to be applied as a substitute for true experimental research to test hypotheses about cause-and-effect relationships in situations in which it is not possible or ethically acceptable to manipulate as in the above mentioned papers, readers should be caution with these cause-and-effect relationships established (Thomas et al., 2015).

Other important limitations of the intervention studies were related to the evaluation. The first weakness was associated with the selection of the test used. Physical fitness can be valid and accurately measured through laboratory tests (Baumgartner et al., 2015). Nevertheless, since laboratory testing requires sophisticated and expensive equipment, qualified examiners, and long testing sessions, this technique is not feasible in PE setting (Meredith & Welk, 2010). Therefore, in the present studies physical fitness field tests with a demonstrated acceptable validity among school age children were used instead. When practical reasons allow doing it, future research studies could examine the effect of similar intervention programs at least by objective instruments such as portable gas analyzers, goniometers, or dynamometers.

A second limitation of the most intervention studies was the fact of not performing a blind evaluation. Due to the restricted human resources available for the most studies, a blind evaluation was not simply possible. Nevertheless, several other measures were always taken into account in order to avoid extraneous variables. For instance, each evaluation was carried out by the same tester, using the same instruments and under the same conditions. The measurements were taken in an indoor sports facility with a non-slippery floor, under the same environmental conditions, on the same day of the week and at the same time for each student. When practical reasons allow it, future research studies should evaluate the effectives of these PE-based interventions through a single-blind cluster randomized controlled trial design.

The last important limitation in the studies with maintenance programs was the fact that the post-detraining values were not measured. Although physical fitness improvements are expected to decrease after a period of detraining (Kenney, Wilmore, & Costill, 2015), a PE-based planning limitation is the fact that the academic year is frequently interrupted by several holiday periods, excursions, and other organized educational activities (Viciana et al., 2014). For instance, the experimental group students performed the development programs during the first semester and then, after a period of detraining coinciding with the Christmas holidays, the students completed the maintenance program during the second semester. Unfortunately, in the current studies the detraining effect previous to the maintenance program could not be examined for practical reasons. Therefore, future research studies should evaluate the physical fitness levels previous to the maintenance program and/or use a second experimental group who after the development program the maintenance program is not applied.

Generally, it would be beneficial for PE teachers to know how to carry out development intervention programs that provide authentic outcomes, then how long it takes to lose the improvements achieved after such a development program, and how a maintenance intervention program should be applied (Viciana, et al., 2014). Consequently, in order to support evidence-based practice in PE for planning health-related physical fitness, a deep examination of many additional related issues should be studied. For instance, for the development programs future research studies should compare the effect of long-term intervention programs with one and two sessions per week, as well as the research into the effectiveness of intervention programs with other frequencies and volume per session is also required. Future research studies should examine if PE-based intervention programs with other frequency, duration and/or type would be also effective in students with high physical fitness baseline levels. Examining in depth the effect of PE-based physical fitness programs on additional health-related psychological markers is also required.

On the other hand, future research interventions should continue examining the effect of different detraining periods among schoolchildren. For example, knowing in depth the detraining effects of each particular health-related physical fitness component depending on the characteristics of the previous development intervention programs is necessary. Finally, regarding the maintenance programs among schoolchildren, future research interventions should be also performed. For instance, since some previous studies have observed statistically significant loss of physical fitness levels after few weeks of detraining (e.g. Ingle et al., 2006; Tsolakis et al., 2004), future research interventions should also examine the effectiveness of interventions that encourage students to perform exercise outside the PE setting for maintaining students´ physical fitness levels. The effectiveness of programs such as irregular TUs in a collaborative project together with students´ families could be examined, especially during long holidays such as Christmas or summer. The effectiveness of maintenance programs consisting of irregular TUs during the school recess and/or extra-curricular time could be also examined. Additionally, future research studies should examine in depth the effects of the maintenance intervention programs with different combinations of frequency, duration and/or type in order to retain the physical fitness gains obtained previously. All of this knowledge could help and guide teachers to design programs that allow a feasible and effective development and maintenance of students' health-related physical fitness in the PE setting.

CONCLUSIONES/

CONCLUSIONS

CONCLUSIONES

General

Los resultados de la presente Tesis Doctoral subrayan que un programa de desarrollo a corto plazo en EF seguido por un programa de mantenimiento es efectivo para incrementar y mantener la condición física relacionada con la salud en escolares. Además del desarrollo y mantenimiento de los niveles de condición física relacionada con la salud, estos programas de intervención permitirían el desarrollo normal de otros contenidos curriculares de EF. Sin embargo, después de un programa de desarrollo a corto plazo en EF, sólo los estudiantes con menores niveles de condición física mejoran su condición física objetiva. Por tanto, con el objetivo real de incrementar la condición física de todos los jóvenes, parece necesario aumentar la carga lectiva de la asignatura de EF. Este conocimiento podría ayudar a los profesores a diseñar programas que permitan un desarrollo y mantenimiento viable y efectivo de la condición física relacionada con la salud en el contexto de la EF escolar.

Específicas

Las conclusiones específicas de la presente Tesis Doctoral fueron las siguientes:

I. Aunque las actitudes y relaciones socio-culturales representan la dimensión prioritaria, la condición física relacionada con la salud sigue teniendo un papel muy importante en el currículo de Real Decreto de Enseñanzas Mínimas en Educación Primaria de la LOE.

II. Los profesores en formación inicial coinciden con el primer nivel curricular de EF en dotar de gran importancia a la condición física relacionada con la salud. Sin embargo, existen importantes deficiencias en su planificación: (a) aplican un número insuficiente de sesiones de intervención que imposibilitaría el incremento significativo de la condición física relacionada con la salud de los escolares requerido por el currículo nacional; y (b) el número de profesores en formación inicial que utiliza refuerzos intermitentes es realmente bajo.

III. Basado en la innovación de una distribución novedosa del tiempo curricular de la EF, las propuestas de unidades didácticas intermitentes, alternadas, irregulares y reforzadas podrían ayudar a una planificación más eficaz de la EF, favoreciendo la intra-disciplinariedad, la significatividad y el mantenimiento del aprendizaje. Este nuevo concepto de unidad didáctica es especialmente importante para ayudar a los profesores de EF a desarrollar y mantener los niveles de condición física relacionada con la salud de los escolares.

IV. Los tests de campo *classic sit-and-reach/ stand-and-reach* y *20-m shuttle run/ 1.5 mile walk/run/ 12 min walk/run* presentan una adecuada validez de criterio para estimar la extensibilidad isquiosural y capacidad cardiorrespiratoria, respectivamente. El sexo, edad y nivel de condición física de los individuos parecen no afectar la validez de criterio de los tests de campo estudiados, a excepción del test *20-m shuttle run* en el que es significativamente mayor para los adultos que para los niños.

V. Después de un programa de acondicionamiento físico en EF, sólo los estudiantes con menor nivel de condición física incrementan la condición física relacionada con la salud. Por tanto, con el objetivo real de incrementar la condición física de todos los jóvenes, parece necesario aumentar la carga lectiva de la asignatura de EF.

VI. Un programa de estiramientos a largo plazo de una sesión semanal mejora la extensibilidad isquiosural de los estudiantes. Un programa de estiramientos a corto plazo realizado una y dos veces por semana mejora la extensibilidad isquiosural de los escolares de forma similar.

VII. Aunque un programa de estiramientos de un minuto por sesión mejora la extensibilidad isquiosural de los escolares, tras un periodo de desentrenamiento de cinco semanas, el nivel de los alumnos vuelve a su valor basal. Sin embargo, a pesar del hecho de que después de un programa de estiramientos de seis minutos por sesión seguido por un periodo de desentrenamiento de cinco semanas los niños también pierden una parte importante de las ganancias obtenidas, sus niveles de extensibilidad isquiosural no vuelven al valor basal.

VIII. Un programa de desarrollo a corto plazo en EF seguido por un programa de mantenimiento es efectivo para incrementar y mantener la condición física relacionada con la salud en escolares. Aunque realizar un refuerzo intermitente consistente en sesiones de acondicionamiento físico alternadas con sesiones de juegos deportivos retiene los niveles de condición física previamente alcanzados, este tipo de programas de mantenimiento implican consumir demasiado tiempo disponible para la planificación, interfiriendo en el desarrollo normal de otros contenidos curriculares en EF. Por el contrario, un refuerzo intermitente que consiste en la realización de actividades deportivas de entrenamiento integrado durante sólo 10-15 minutos seguido por actividades de aprendizaje de estos deportes durante dos sesiones por semana, no sólo es eficaz sino también viable en el mantenimiento de los niveles de fuerza muscular y capacidad cardiorespiratoria en el contexto de la EF escolar. De manera similar, después de un programa de desarrollo de la flexibilidad a corto plazo, un refuerzo intermitente con dos sesiones semanales de sólo un minuto también es viable y eficaz para el mantenimiento de la extensibilidad isquiosural.

IX. Aunque un programa de acondicionamiento físico en EF mejora la condición física objetiva en escolares, este tipo de intervenciones no cambia la condición física percibida de los estudiantes. La mejora del programa de acondicionamiento físico en EF sobre la condición física objetiva tampoco va acompañada de grandes cambios en el autoconcepto físico, a pesar de que el programa de intervención parece mantener la percepción de los escolares sobre su apariencia física, fuerza muscular y autoestima.

X. Un programa de acondicionamiento físico en EF solo mejora los niveles de capacidad cardiorrespiratoria de los estudiantes con una moderada-alta motivación autodeterminada hacia la EF. Por tanto, con el fin de aumentar la condición física relacionada con la salud de sus estudiantes, los profesores de EF deberían promover la motivación hacia la EF.

CONCLUSIONS

Overall

The results of the current Doctoral Thesis underline that a short-term PE-based development program followed by a maintenance program is effective in increasing and maintaining objective health-related physical fitness in schoolchildren. In addition to the development and maintenance of students´ health-related physical fitness levels, these intervention programs would permit the regular development of other PE curricular contents. However, after a short-term PE-based development program, only students with lower physical fitness levels improve objective health-related physical fitness. Therefore, in order to improve physical fitness of all young people, it seems necessary to increase the academic load of the PE subject. This knowledge could help teachers to design programs that permit a feasible and effective development and maintenance of health-related physical fitness in the PE setting.

Specifics

The specific conclusions of the present Doctoral Thesis were the following:

I. Although the attitudes and socio-cultural relations represent the priority dimension, health-related physical fitness still has a very important role in the standard curriculum of Royal Decree Minimum Teaching in Primary Education of LOE.

II. Pre-service teachers coincide with the PE standard curriculum on providing great importance to the health-related physical fitness. However, there are important deficiencies in their planning: (a) they apply an insufficient number of intervention sessions that would make impossible the significant improvement of health-related physical fitness in schoolchildren required by the standard curriculum; and (b) the number of pre-service teachers using intermittent reinforcements is really low.

III. Based on the innovation of a new distribution of PE curriculum time, the proposals of intermittent, alternated, irregular, and reinforced TUs could help in a more effective PE planning, promoting the intra-disciplinary, the significance and the learning maintaining. This new concept of TU is especially important to help PE teachers to develop and maintain schoolchildren´s health-related physical fitness levels.

IV. The field tests classic sit-and-reach/ stand-and-reach and 20-m shuttle run/ 1.5 mile walk/run/ 12 min walk/run show an adequate criterion-related validity for estimating hamstring extensibility and cardiorespiratory fitness, respectively. Sex, age and fitness level of individuals seem not to affect the criterion-related validity of the studied field tests, except for the 20-m shuttle run test in which is significantly higher for adults than for children.

V. After a PE-based development program, only students with lower physical fitness levels improve health-related physical fitness. Therefore, in order to improve physical fitness of all young people, it seems necessary to increase the academic load of the PE subject.

VI. A long-term stretching program performed once a week improves students´ hamstring extensibility. A short-term stretching program performed once and twice a week improves students´ hamstring extensibility in a similar manner.

VII. Although a one-minute stretching program improves schoolchildren’s hamstring extensibility, after five-week detraining period, students’ level reverts back to its baseline. However, despite the fact that after a six-minute stretching program followed by five-week detraining period children also lose a significant part of the obtained gains, their hamstring extensibility levels do not revert to baseline.

VIII. A short-term PE-based development program followed by a maintenance program is effective in increasing and maintaining health-related physical fitness in schoolchildren. Although carrying out an intermittent reinforcement consisting of a physical fitness session alternated with a sport games session retains the physical fitness levels previously achieved, this kind of maintenance program involve consuming too much time available for planning, interfering with the normal development of other curricular content in PE. On the contrary, an intermittent reinforcement consisting of performing sports-integrated training activities for only 10-15 min followed by introduction activities to these sports two sessions per week, it is not only effective but also feasible for maintaining schoolchildren´s muscular and cardiorespiratory fitness levels in PE setting. Similarly, after a short-term stretching development program, an intermittent reinforcement with only one-minute sessions twice a week is also feasible and effective in maintaining students´ hamstring extensibility.

IX. Although a PE-based physical fitness program improves objective health-related physical fitness in schoolchildren, this kind of interventions does not change students´ perceived physical fitness. The PE-based physical fitness program improvement in objective physical fitness is not either accompanied by major changes in physical self-concept, even though the intervention program seems to maintain the schoolchildren´s perceptions in physical appearance, muscular strength, and self-esteem.

X. A PE-based physical fitness program only improves the cardiorespiratory fitness levels of students with a moderate-to-high self-determined motivation toward PE. Therefore, in order to increase the health-related physical fitness of their students, PE teachers should promote motivation toward PE.

REFERENCES

[REFERENCIAS BIBLIOGRÁFICAS]

REFERENCES [REFERENCIAS BIBLIOGRÁFICAS]

Alfermann, D., & Stoll, O. (2000). Effects of physical exercise on self-concept and well-being. *International Journal of Sport Psychology*, *31*, 47-65. doi: 10.1080/1061580021000020752

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*, 217-238.

Baumgartner, T. A., Jackson, A. S., Mahar, M. T., & Rowe, D. A. (2015). *Measurement for evaluation in kinesiology* (9th ed.). Burlington: Jones & Bartlett Learning.

Carraro, A., Scarpa, S., & Ventura, L. (2010). Relationships between physical self-concept and physical fitness in Italian adolescents. *Perceptual and Motor Skills*, *110*, 522-530.

Castro-Piñero, J., Artero, E. G., España-Romero, V., Ortega, F. B., Sjöström, M., & Ruiz, J. R. (2010). Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine*, *44*, 934-943. doi: 10.1136/bjsm.2009.058321

Chatterjee, P., Banerjee, A. K., & Majumdar, P. (2006). Validity of the 20-m multi stage shuttle run test for the prediction of VO_2max in junior taekwondo players of India. *International Journal of Sports Science*, *18*, 1-7.

Cooper, H., Hedges, L. V, & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Sage Publications.

Cuenca-García, M., Ortega, F. B., Ruiz, J. R., González-Gross, M., Labayen, I., Jago, R… Sjöström, M. (2014). Combined influence of healthy diet and active lifestyle on cardiovascular disease risk factors in adolescents. *Scandinavian Journal of Medicine & Science in Sports*, *24*, 553-562. doi: 10.1111/sms.12022.

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227-268. doi: 10.1207/S15327965PLI1104_01

Emmer, E. T. (1986). Academic activities and task in first-year teachers’ classes. *Teaching and Teacher Education*, *2*, 229-244. doi: 10.1016/S0742-051X(86)80004-X

European Commission/ EACEA/ Eurydice (2013). *Physical education and sport at school in Europe Eurydice Report*. Luxembourg: Publications Office of the European Union.

Faigenbaum, A. D., Bush, J. A., McLoone, R. P., Kreckel, M. C., Farrell, A., Ratamess, N. A., & Kang, J. (2015). Benefits of strength and skill-based training during primary school physical education. *Journal of Strength and Conditioning Research*, *29*, 1255-1262. doi: 10.1519/JSC.0000000000000812.

Faigenbaum, A., Zaichkowsky, L. D., Westcott, W. L., Long, C. J., LaRosa-Loud, R., Micheli, L. J., & Outerbridge, A. R. (1997). Psychological effects of strength training on children. *Journal of Sport Behavior*, *20*, 164-175.

Fairclough, S., & Stratton, G. (2005). Physical activity levels in middle and high school physical education: A review. *Pediatric Exercise Science*, *17*, 217-236.

Fairclough, S., & Stratton, G. (2006). A review of physical activity levels during elementary school physical education. *Journal of Teaching in Physical Education*, *25*, 239-257.

Feldman, D. E., Shrier, I., Rossignol, M., & Abenhaim, L. (2001). Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology*, *154*, 30-36.

Greene, J. C., & Ignico, A. A. (1995). The effects of a ten-week physical fitness program on fitness profiles, self-concept, and body-esteem in children. *Journal of the International Council for Health, Physical Education, Recreation, Sport, and Dance*, *31*, 42-47.

Hardman, K. (2008). Physical education in schools: A global perspective. *Kinesiology*, *40*, 5-28.

Harter, S. (2012). *The construction of self. Developmental and sociocultural foundations* (2nd ed.). London: Guilford Press.

Ingle, L., Sleap, M., & Tolfrey, K. (2006). The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *Journal of Sports Sciences*, *24*, 987-997.

Jefatura del Estado (1985). Ley Orgánica 8/1985, de 3 de julio, Reguladora del Derecho a la Educación. *Boletín Oficial del Estado*, *159*, 21015-21022.

Jefatura del Estado (1990). Ley Orgánica 1/1990 de 3 de Octubre, de Ordenación General del Sistema Educativo. *Boletín Oficial del Estado*, *238*, 28927-28942.

Jefatura del Estado (2006). Ley Orgánica 2/2006 de 3 de Mayo, de Educación. *Boletín Oficial del Estado*, *106*, 17158-17207.

Jefatura del Estado (2013). Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa. *Boletín Oficial del Estado*, *295*, 97858-97921.

Jones, M. A., Stratton, G., Reilly, T., & Unnithan, V. (2005). Biological risk indicators for recurrent non-specific low back pain in adolescents. *British Journal of Sports Medicine*, *39*, 137-140. doi: 10.1136/bjsm.2003.009951

Kajala, S., Jaakkola, T., & Liukkonen, J. (2010). Fundamental movement skills and motivational factors influencing engagement in physical activity. *Perceptual & Motor Skills*, *111*, 115-128. doi: 10.2466/06.10.25.PMS.111.4.115-128

Kelly, L. E., & Melograno, V. J. (2004). *Developing the physical education curriculum. An achievement-based approach*. Champaign: Human Kinetics.

Kenney, W. L., Wilmore, J. H., & Costill, D. L. (2015). *Physiology of sport and exercise* (6th ed.). USA: Human Kinetics.

Klesges, G., Haddock, C. K., Stein, R. J., Klesges, L. M., Eck, L. H., & Hanson, C. L. (1992). Relationship between psychosocial functioning and body fat in preschool children: A longitudinal investigation. *Journal of Consulting & Clinical Psychology*, *60*, 793-796.

Kujala, U. M., Taimela, S., Salminen, J. J., & Oksanen, A. (1994). Baseline anthropometry, flexibility and strength characteristics and future low-back pain in adolescent athletes and nonathletes. *Scandinavian Journal of Medicine and Science in Sports*, *4*, 200-205. doi: 10.1111/j.1600-0838.1994.tb00426.x

Lonsdale, C., Sabiston, C. M., Raedeke, T. D., Ha, A. S. C., & Sum, R. K. W. (2009). Self-determined motivation and students´ physical activity during structured physical education lessons and free choice periods. *Preventive Medicine*, *48*, 69-73. doi: 10.1016/j.ypmed.2008.09.013

Mayorga-Vega, D., & Viciana, J. (2014). Adolescents´ physical activity in physical education, school recess, and extra-curricular sport by motivational profiles. *Perceptual and Motor Skills*, *118*, 663-679. doi: 10.2466/06.10.PMS.118k26w0

Mayorga, D., Viciana, J., & Cocca, A. (2012). Relationship between physical self-concept and health-related physical fitness in Spanish schoolchildren. *Procedia. Social and Behavioral Sciences*, *69*, 659-668. doi: 10.1016/j.sbspro.2012.11.458

Meredith, M. D., & Welk, G. J. (Eds.) (2010). *FITNESSGRAM® & ACTIVITYGRAM® test administration manual* (4th ed.). Champaign: Human Kinetics.

Meyer, U., Schindler, C., Zahner, L., Ernst, D., Hebestreit, H., van Mechelen, W… Kriemler, S. (2014). Long-term effect of a school-based physical activity program (KISS) on fitness and adiposity in children: A cluster-randomized controlled trial. *Plos One*, *9*, e87929. doi: 10.1371/journal.pone.0087929

Mikkelsson, L., Nupponen, H., Kaprio, J., Kautiainen, H., Mikkelsson, M., & Kujala, U. M. (2006). Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine*, *40*, 107-113.

Ministerio de Educación, Cultura y Deporte (2014). 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*, 19349-19420.

Ministerio de Educación, Cultura y Deporte (2015). Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. *Boletín Oficial del Estado*, *3*, 169-546.

Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology*, *71*, 225-242. doi: 10.1348/000709901158497

Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity* , *32*, 1-11.

Ortega, F. B., Ruiz, J. R., Castillo, M. J., Moreno, L. A., González-Gross, M., Wärnberg, J., & Gutiérrez, A. (2005). Low level of physical fitness in Spanish adolescents. Relevance for future cardiovascular health (AVENA study). *Revista Española de Cardiología*, *58*, 898-909.

Pérez-Pastur, R. (2010). Unidad didáctica sobre las capacidades físicas básicas [Teaching unit on the basic physical abilities]. *Eduinnova*, *25*, 95-106.

Pescatello, L. S., Arena, R., Riebe, D., & Thompson, P. D. (2014). *ACSM's guidelines for exercise testing and prescription* (9th ed.). Philadelphia: Wolters Kluwer/ Lippincott Williams & Wilkins.

Placek, J. H. (1984). A multi-case study of teacher planning in physical education. *Journal of Teaching in Physical Education*, *4*, 39-49.

Planinsec, J., & Fosnaric, S. (2005). Relationship of perceived physical self-concept and physical activity level and sex among young children. *Perceptual and Motor Skills*, *100*, 349-353.

Ramírez, J., Muros, J. J., Morente, J., Sánchez, C., Femia, P., & Zabala, M. (2012). Efecto de un programa de entrenamiento aeróbico de 8 semanas durante las clases de educación física en adolescentes [Effect of an 8-week aerobic training program during physical education lessons on aerobic fitness in adolescents]. *Nutrición Hospitalaria*, *27*, 747-754. doi: 10.3305/nh.2012.27.3.5725

Robles, J., Giménez, F. J., & Abad, M. T. (2010). Motivos que llevan a los profesores de educación física a elegir los contenidos deportivos en la E.S.O. [Motives for physical education teachers to select sport contents in secondary school]. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, *18*, 5-8.

Ruiz, J. R., Castro-Pinero, J., Artero, E. G., Ortega, F. B., Sjöström, 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. doi: 10.1136/bjsm.2008.056499

Ruiz, J. R., Ortega, F. B., & Castro-Piñero, J. (2015). Validity and reliability of the 1/4 mile run-walk test in physically active children and adolescents. *Nutrición Hospitalaria*, *31*, 875-882. doi: 10.3305/nh.2015.31.2.7789

Ryan, R. M., Williams, G. C., Patrick, H., & Deci, E. L. (2009). Self-determination theory and physical activity: The dynamics of motivation in development and wellness. *Hellenic Journal of Psychology*, *6*, 107-124.

Sadres, E., Eliakim, A., Constantini, N., Lidor, R., & Falk, B. (2001). The effect of long-term resistance training on anthropometric measures, muscle strength, and self concept in pre-pubertal boys. *Pediatric Exercise Science*, *13*, 357-372.

Salokun, S. O. (1994). Positive change in self-concept as a function of improved performance in sports. *Perceptual and Motor Skills*, *78*, 752-754.

Sánchez, E., Mayorga-Vega, D., Fernández, E., & Merino-Marban, R. (2014). Efecto de un programa de estiramiento de la musculatura isquiosural en las clases de educación física en educación primaria [Effect of a hamstring stretching programme during physical education lessons in primary education]. *Journal of Sport and Health Research*, *6*, 159-168.

Siedentop, D., & Tanehill, D. (2000). *Developing teaching skills in physical education* (4th ed.). Mountain View: Mayfield Publishing Company.

Sjölie, A. N. (2004). Low-back pain in adolescents is associated with poor hip mobility and high body mass index. *Scandinavian Journal of Medicine and Science in Sports*, *14*, 168-175.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). California: Sage Publications.

Smith, W. W. (2011). Skills acquisition in physical education: A speculative perspective. *Quest, 63*, 265-274.

Sygit, K., Kołłątaj, W., Goździewska, M., Sygit, M., Kołłątaj, B., & Karwat, I. D. (2012). Lifestyle as an important factor in control of overweight and obesity among schoolchildren from the rural environment. *Annals of Agricultural and Environmental Medicine*, *19*, 557-561.

Tomás, I., Marsh, H. W., González-Romá, V., Valls, V., & Nagengast, B. (2014). Testing measurement invariance across Spanish and English versions of the physical self-description questionnaire: An application of exploratory structural equation modeling. *Journal of Sport & Exercise Psychology*; *36*, 179-188. doi: 10.1123/jsep.2013-0070.

Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2015). *Research methods in physical activity* (7th ed.). Champaign: Human Kinetics.

Tsolakis, C. K., Vagenas, G. K., & Dessypris, A. G. (2004). Strength adaptations and hormonal responses to resistance training and detraining in preadolescent males. *Journal of Strength and Conditioning Research*, *18*, 625-629.

Ullrich-French, S., & Cox, A. (2009). Using cluster analysis to examine the combinations of motivation regulations of physical education students. *Journal of Sport and Exercise Psychology*, *31*, 358-379.

Van der Mars, H. (2006). Time and learning in physical education. In D. Kirk, D. Mcdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 191-213). London: Sage Publications.

Viciana, J., Mayorga-Vega, D., & Merino-Marban, R. (2014). Physical education-based planning for developing and maintaining students' health-related physical fitness levels. In R. Todaro (Ed.), *Handbook of physical education research. Role of school programs, children´s attitudes and health implications* (pp. 237-252). New York: Nova science Publisher.

Viciana, J. (2002). *Planificar en Educación Física* [Planning in Physical Education]. Barcelona: INDE.

Viciana, J., Salinas, F., & Cocca, A. (2007). Análisis de contenido comparativo del primer nivel curricular de Educación Física en Secundaria [Analysis of comparative content of the first level curricular of Physical Education in Secondary]. *Profesorado*, *11*, 1-13.

Viciana, J., & Zabala, M. (2004). Un estudio descriptivo sobre cómo planifican los profesores de Educación Física [A descriptive study about how to physical education teachers plan]. In M. González, A. Sánchez, & J. Gómez (Eds.) *International Congress of AIESEP 2002. Preparación profesional y necesidades sociales* (pp. 732-739). A Coruña: AIESEP.

Von Haaren, B., Härtel, S., Seidel, I., Schlenker, L., & Bös, K. (2011). Die validität des 6-minuten-laufs und 20m shuttle runs bei 9- bis 11-jährigen kindern [Validity of a 6-min endurance run and a 20-m shuttle run in 9-to 11-year old children]. *Deutsche Zeitschrift für Sportmedizin*, *62*, 351-355.

World Health Organization (1946). *Constitution of the World Health Organization*. New York: International Health Conference.

Yli-Piipari, S., Watt, A., Jaakkola, T., Liukkonen, J., & Nurmi, J. (2009). Relationship between physical education students' motivational profiles, enjoyment, state anxiety, and self-reported physical activity. *Journal of Sports Science & Medicine*, *8*, 327-336.

Zhu, W., Mahar, M. T., Welk, G. J., Going, S. B., & Cureton, K. J. (2011). Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Journal of Preventive Medicine*, *41*, S68-76. doi: 10.1016/j.amepre.2011.07.001

ANEXO

[ANNEX]

ANEXO [ANNEX]

1. Carta de aprobación del protocolo del estudio por el Comité de Ética de la Universidad de Granada.

Universidad de Granada
Vicerrectorado de Política
Científica e Investigación

COMISIÓN DE ÉTICA EN INVESTIGACIÓN
DE LA UNIVERSIDAD DE GRANADA

La Comisión de Ética en Investigación de la Universidad de Granada, oído el informe preliminar del Presidente del Comité en Investigación Humana, emite informe favorable a la metodología en la investigación titulada "EFECTOS DEL REFUERZO INTERMITENTE SOBRE EL MANTENIEMIENTO DE LA CONDICION FISICA ORIENTADA A LA SALUD EN LA EDUCACION FISICA ESCOLAR" que dirige D./Dña. Daniel Mayorga Vega, quedando registrada con el nº: 753.

Granada a 14 de noviembre de 2012

LA PRESIDENTA

Fdo: Mª Dolores Suárez Ortega

LA SECRETARIA

Fdo: Irene Luque Fernández

AGRADECIMIENTOS

[ACKNOWLEDGMENTS]

AGRADECIMIENTOS [ACKNOWLEDGMENTS]

Una vez llegado a este punto es cuando uno entiende que la presente Tesis Doctoral ha sido posible gracias no solo a la contribución de grandes profesionales, sino también de grandes personas que de una u otra manera me han marcado a lo largo de toda mi vida. Por ello, no podría acabar esta tesis sin agradecerles a todos ellos lo mucho que me han dado:

A mi director de tesis, Jesús, por confiar en mí desde el primer momento. Por compartir toda tu sabiduría desinteresadamente. Por estar siempre disponible fuera cuando fuera… ahí tú estabas cada vez que te necesitaba. Por soportar con paciencia esas eternas tutorías, especialmente en aquellos duros inicios. Porque siempre supiste motivarme cuando más lo necesitaba. Por trasmitirme tu pasión por la investigación, por ser un modelo a seguir... Porque por encima de un excelente director, has sido un gran amigo… sin ti esta tesis no hubiese sido posible.

Al que fue mi profesor y ahora es un gran amigo, Rafa, quien primero confió en mí y probablemente el gran culpable de todo esto…. A ti que me supiste emocionar con la investigación. Sin ese primer paso no habría un camino… Además sin tu gran ayuda y contribución indudablemente esta tesis tampoco hubiese sido la misma. *To my tutor in the research stay, Janusz, who lighted me with all his knowledge and who did my research stay an unforgettable and fruitful experience. I wish I can continue sharing good moments with you.* A Humberto quien con paciencia y cariño siempre me ilustraba con sus conocimientos de estadística. A los numerosos editores y revisores anónimos que a lo largo de este camino me enseñaron y ayudaron a mejorar cada día. A los numerosos profesores y autores, que compartieron su conocimiento conmigo. Porque cada granito de arena es necesario para formar una montaña, todo lo que sé se lo tengo que agradecer a cada uno de ellos.

A Aliisa y Ania, que con gran paciencia y dedicación, corregían nuestros borradores. Aunque los inicios con el inglés fueron duros, ellas siempre estuvieron ahí para ayudarnos… ¡Jesús que haríamos nosotros sin ellas! A Adrián que de forma desinteresada me dio toda su ayuda cuando más lo necesitaba. Te estoy infinitamente agradecido, sin ti este proyecto tampoco habría sido posible. A los bibliotecarios de nuestra facultad, en especial a Ana, por su gran ayuda y dedicación.

Como no podría ser de otra manera también le doy gracias a todos los alumnos y padres por su participación voluntaria e incondicional en los estudios. Agradezco a todos los miembros del centro escolar por su entusiasmo y colaboración, ellos siempre me abrieron las puertas. Un especial agradecimiento a los profesores de educación física Jorge, Moisés y Carmen, por compartir esta experiencia tan enriquecedora conmigo. Ellos estuvieron ahí en cada momento que se les necesitaba.

A toda mi familia, en especial a mi abuela María y mi tía Reme, que con paciencia y dedicación también me educasteis y cuidasteis en mi infancia… Un especial agradecimiento a mis primos Samu y Migue, por todos aquellos buenos momentos que pasamos en nuestra niñez… Aunque el tiempo pasa deprisa, estamos hechos de recuerdos y esos marcaron mi vida… *Dla calej rodziny z Polski ktorzy przyjeli mnie do siebie od pierwszego dnia. Maciek dla ciebie za to ze czynisz kazde moje wakacje niezapomniane i traktujesz mnie jak brata*

Al amor de mi vida, Ania, que desde el primer momento me hiciste sentir diferente… Tú que desde el primer día me lo has dado todo. A ti que has sufrido con gran paciencia todo el trabajo que supone esta tesis, que siempre has estado apoyándome sin descanso…. A ti que siempre supiste darme ánimos cuando más lo necesitaba. Tu fortaleza ha sido todo un modelo para mí. Y aunque a veces en la vida las cosas sean duras, nunca olvides que juntos superaremos cada dificultad que nos venga…. *Kocham cię!*

A mi hermana, Lidia, por darme tan buenos momentos. Crecer contigo fue una experiencia inolvidable y me regalaste la mejor niñez que se podía tener. A ti porque cuando se te necesita siempre estás ahí…

A mis padres, Alfonso y María Dolores, ¿cómo agradeceros a vosotros con tan solo unas palabras? Con el fruto de vuestro amor me disteis la vida. Con gran esfuerzo, paciencia y dedicación me disteis una buena educación. Hicisteis que los años de mi niñez fueran inmejorables. Durante la juventud me aconsejasteis y apoyasteis ante cada dedición y ante cada dificultad. Y no conformes con todo ello, como si ya no hubieseis hecho suficiente, casi 30 años más tarde seguís ahí dándomelo todo. Y por todo ello siempre os estaré eternamente agradecido.

Tesis Doctoral Internacional
Programa Oficial de Doctorado en Ciencias de la Educación

Departamento de Educación Física y Deportiva
Facultad de Ciencias del Deporte
Universidad de Granada