

Validade e reprodutibilidade dos instrumentos de medida da atividade física do tipo *self-report* em adolescentes: uma revisão sistemática

Validity and reliability of self-report instruments for measuring physical activity in adolescents: a systematic review

José Cazuza de Farias Júnior ¹
Adair da Silva Lopes ²
Alex Antonio Florindo ³
Pedro C. Hallal ^{4,5}

Abstract

This was a systematic review of studies on the reliability and validity of self-report instruments for measuring physical activity, or subjective measurements, in adolescents (10-18 years). Searches were conducted in databases (MEDLINE, PsycInfo, SportsDiscus, Scopus, Web of Science, SciELO, Lilacs) and in the references of the retrieved articles. Sixty-six studies met the inclusion criteria. The majority were from North America, with only 5 from Brazil. Fifty-two different instruments were identified: 42 questionnaires, 6 diaries or logs, and 4 interviews. "Test-retest" reliability varied from 0.20 to 0.98; the majority (28/50) of the coefficients showed values < 0.70. Validity coefficients showed wide variation (-0.13 to 0.88), with the majority (64/84) ≤ 0.50. Only 3 instruments displayed correlations ≥ 0.70. Various instruments were tested in adolescents, especially questionnaires. These instruments generally showed better "test-retest" reliability than validity.

Motor Activity; Adolescent; Reproducibility of Results; Validity of Tests; Methods

Introdução

Há um crescente reconhecimento da importância da avaliação do nível de prática de atividade física em estudos epidemiológicos ^{1,2}. Entretanto, obter dados precisos sobre o nível de atividade física, sobretudo em crianças e adolescentes, é um grande desafio em razão da complexidade que caracteriza tal comportamento nesse grupo populacional ^{3,4,5}.

A atividade física representa qualquer movimento corporal produzido pelos músculos esqueléticos que resulta em gasto energético acima dos níveis de repouso (lazer, trabalho, deslocamentos, atividades do lar) ^{6,7}. Esse conceito tem sido unanimemente aceito e amplamente adotado por pesquisadores de diversas áreas, em particular da área da saúde. O exercício físico é um tipo de atividade física, caracterizada por ser planejada, estruturada e repetitiva, que objetiva melhorar e/ou manter um ou mais componentes da aptidão física ⁷.

Em estudos com o objetivo de estimar a prevalência de atividade física, identificar fatores associados, analisar a sua relação com diferentes desfechos de interesse e avaliar a efetividade de programas de intervenção é essencial dispor de instrumentos de medida da atividade física que apresentem propriedades psicométricas adequadas (reprodutibilidade e validade) ^{4,5,8}.

Há uma grande diversidade de métodos e técnicas para mensurar a atividade física, sen-

¹ Departamento de Educação Física, Universidade Federal da Paraíba, João Pessoa, Brasil.

² Centro de Desportos, Universidade Federal de Santa Catarina, Florianópolis, Brasil.

³ Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brasil.

⁴ Programa de Pós-graduação em Educação Física, Universidade Federal de Pelotas, Pelotas, Brasil.

⁵ Programa de Pós-graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, Brasil.

Correspondência

J. C. Farias Júnior
Departamento de Educação Física, Universidade Federal da Paraíba.
Campus I, João Pessoa, PB
88010-970, Brasil.
jcazuzajr@hotmail.com

do eles classificados como objetivos e subjetivos 4,5,8. Os métodos objetivos (sensores de movimento, monitores de frequência cardíaca, água duplamente marcada – DLW, observação direta do comportamento), embora apresentem níveis considerados satisfatórios de reprodutibilidade e validade, possuem limitações em termos de logística e custo, podem ser reativos e mensuram apenas atividades físicas “atuais” ou “recentes” 8,9,10,11.

Apesar do crescimento na utilização de acelerômetros e pedômetros para mensurar a atividade física em adolescentes 10,12, em estudos epidemiológicos a medida de atividade física tem sido efetuada, essencialmente, por métodos subjetivos (diários, questionários, entrevistas estruturadas), em particular em países de renda baixa e média 1,2. Isso se deve à praticidade dessas medidas, ao baixo custo, por não serem reativos e permitirem, com algumas variações, mensurar as atividades físicas praticadas em um ou mais domínios, por diferentes períodos de tempo 4,5,8,11.

Vários estudos de revisão sobre métodos de medida da atividade física aplicados em adolescentes foram publicados 4,5,8,10,12,13. Os que analisaram a reprodutibilidade e validade dos instrumentos do tipo *self-report* (medidas subjetivas da atividade física) não recorreram ao método de revisão sistemática 8,9,12. Os estudos que utilizaram esse método foram realizados há quase uma década 8 ou mais 14, ou analisaram as publicações durante um período curto de tempo (estudos publicados durante a década de 1990) 15.

Com intuito de preencher essa lacuna de conhecimento foram analisados, mediante método de revisão sistemática, os estudos de reprodutibilidade e validade dos instrumentos de medida

da atividade física do tipo *self-report* testados em adolescentes (10-19 anos).

Métodos

As buscas foram realizadas em bases eletrônicas de dados, considerando o período de janeiro de 1980 a dezembro de 2007. A Tabela 1 apresenta as informações sobre as bases de dados consultadas, descritores e operadores utilizados na busca. O processo de revisão dos estudos está descrito na Figura 1.

Os seguintes critérios de inclusão foram adotados para esta revisão: (1) ser estudo de reprodutibilidade e/ou validade de instrumentos de medida da atividade física do tipo *self-report*; (2) incluir adolescentes de 10-19 anos ou média de idade dentro desse intervalo; (3) ter utilizado como critério de referência – “padrão-ouro” – ao menos uma medida objetiva e/ou subjetiva de atividade física e/ou de aptidão cardiorrespiratória; (4) ser artigo original de pesquisa desenvolvida com seres humanos, publicado em periódico indexado nas bases utilizadas, em inglês/português/espanhol, durante janeiro de 1980 a dezembro de 2007.

Artigos de revisão, monografias, dissertações, teses, resumos, capítulos ou livros e ponto de vista/opinião de especialistas não foram incluídos. Também foram excluídos os estudos com adolescentes de grupos especiais e aqueles cujas medidas de reprodutibilidade e/ou validade foram apresentadas apenas como parte dos métodos (outros objetivos).

Estudos cujos resumos atenderam aos critérios de inclusão ou não apresentaram informações suficientes para que fossem excluídos de

Tabela 1

Bases de dados eletrônicas, descritores e operadores utilizados nas buscas.

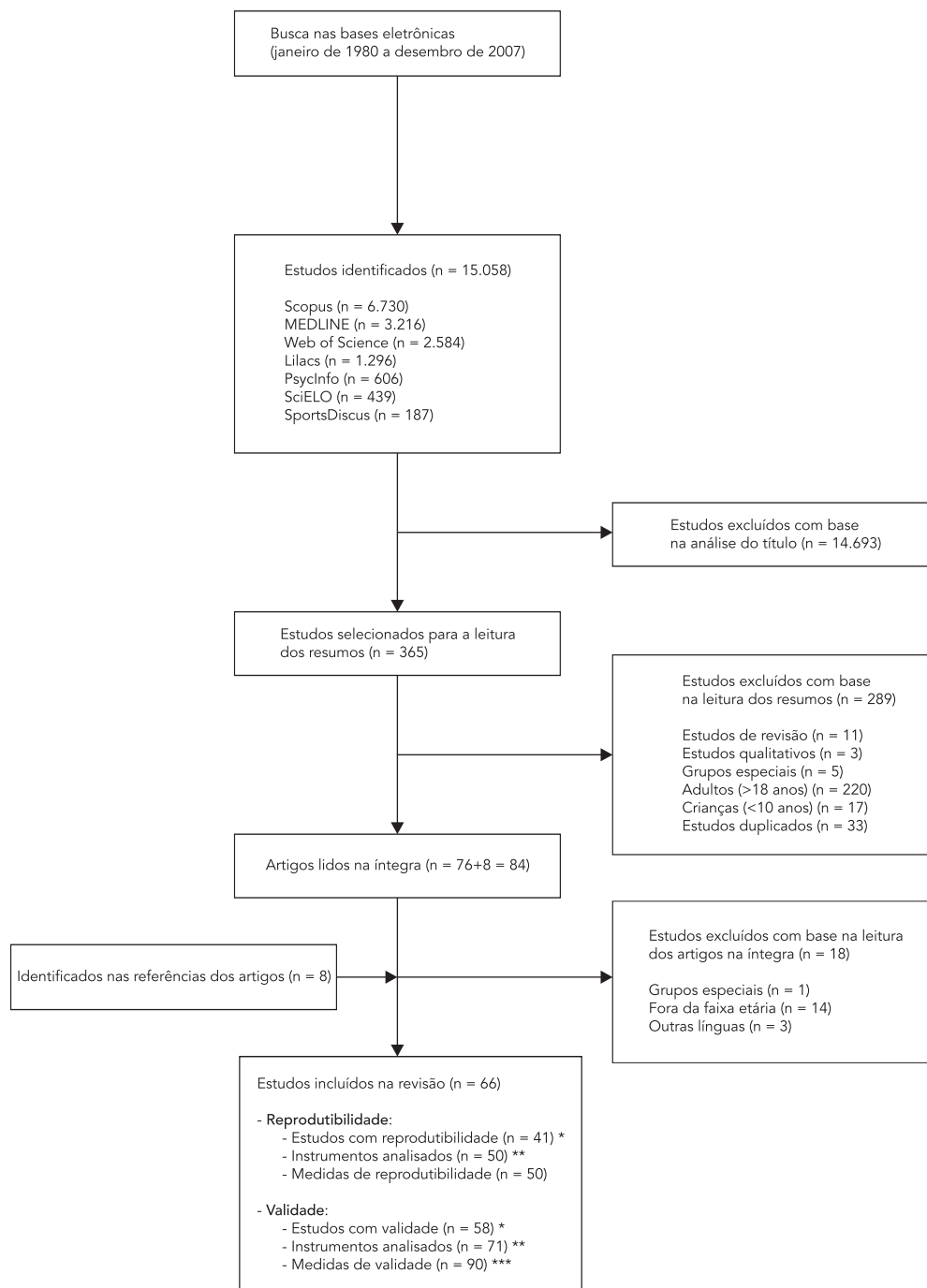
Base de dados	Descritores *	Operadores **
MEDLINE PsycInfo SportsDiscus Lilacs Scopus Web of Science SciELO	<p>1. Instrumento: self-report, checklist, recall, 24h, interviews, questionnaire, diary, assessment, survey, measurement, log.</p> <p>2. Medida psicométrica: reproducibility, test-retest, reliability, validity, validation, psychometric analysis, development.</p> <p>3. Atividade física: physical activity, leisure activities, exercise, motor activity, sports, transportation, working, organized, unorganized, daily physical activity, moderate, vigorous, energy expenditure.</p> <p>4. Grupo populacional: children, child, adolescent, adolescence, youth, teen, teenager, high school, students, young.</p>	And, Or, Not

* Descritores em inglês/português, no título e/ou no resumo do artigo;

** Combinar os descritores e termos utilizados.

Figura 1

Etapas do processo de revisão sistemática.



* Nem todos os estudos analisaram a reprodutibilidade e validade, simultaneamente;

** Alguns estudos analisaram mais de um instrumento (reprodutibilidade n = 6; validade n = 10);

*** Alguns instrumentos foram comparados contra mais de um critério de referência – “padrão-ouro” (n = 15).

forma imediata foram obtidos e lidos na íntegra. Efetuou-se uma revisão nas referências desses artigos com intuito de identificar algum estudo potencialmente relevante que não havia sido identificado na busca eletrônica.

A extração das informações nos artigos selecionados foi efetuada, em duplicata, por um único revisor, que utilizou uma ficha padronizada. As divergências observadas entre as duas revisões foram analisadas por mais dois pesquisadores, de forma independente, e foram resolvidas por meio de consenso. Informações que foram coletadas: (1) o primeiro autor do estudo, ano de publicação, número de sujeitos na amostra, total e por sexo, faixa etária, nome do país onde foi realizado o estudo, procedimento de seleção da amostra; (2) tipo de instrumento (diário, questionário, entrevista estruturada), nome original do instrumento, período de referência da sua medida (exemplo, um dia, semana típica, últimos sete dias, último ano), capacidade do instrumento de realizar estimativas do nível de atividade física de acordo com as recomendações atuais para adolescentes; (3) o critério de referência – “padrão-ouro” – adotado na validação, a análise utilizada para estimar a reprodutibilidade “teste-reteste” e a validade, o intervalo entre as réplicas de aplicação do instrumento – “teste-reteste”; (4) a medida de reprodutibilidade e/ou validade, se o estudo comparou as medidas por sexo e idade dos adolescentes, e intensidade das atividades físicas (moderadas *vs.* vigorosas).

Resultados

Foram identificados 15.066 estudos, considerando os dois métodos de busca (bases eletrônicas $n = 15.058$, referências dos artigos $n = 8$). Sessenta e seis estudos atenderam aos critérios de inclusão. Dentre eles, 41 haviam analisado a reprodutibilidade “teste-reteste” e 58 a validade. Alguns estudos analisaram mais de um instrumento ($n = 10$)^{16,17,18,19,20,21,22,23,24,25} e, nos estudos com análise de validade, 15 instrumentos foram comparados contra mais de um critério de referência – “padrão-ouro”^{16,23,26,27,28,29,30,31,32,33,34,35}. Em função disso, nesta revisão, o “instrumento” será a unidade de análise e, por conseguinte, em algumas vezes os valores de “n” (ou o número de referências citadas) serão diferentes do número de estudos revisados (Figura 1).

A Tabela 2 apresenta uma descrição dos 66 estudos revisados. Foram analisados 79 instrumentos, sendo 16 diários, 57 questionários e 6 entrevistas estruturadas. A análise da reprodutibilidade foi efetuada em 50 instrumentos e da validade em 71. Ou seja, em alguns instrumentos

foi analisada apenas a reprodutibilidade ($n = 8$)^{36,37,38,39,40,41,42,43}, em outros a validade ($n = 29$)^{16,18,21,25,28,30,31,32,33,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60}.

Identificaram-se 52 instrumentos diferentes (42 questionários, 6 diários e 4 entrevistas estruturadas), mas apenas 11 haviam sido testados mais de uma vez^{16,18,20,21,22,24,25,26,29,30,31,33,35,36,37,39,41,44,46,49,52,54,55,58,59,60,61,62,63,64,65,66,67}. A maioria destes instrumentos (33/52)^{19,20,21,22,23,24,25,27,28,32,35,38,40,41,46,47,50,51,53,55,56,57,61,64,68,69,70,71,72,73,74} não permitia realizar estimativas do nível de atividade física dos adolescentes de acordo com as recomendações dos 300 min/semana⁷⁵. Dos que realizavam tais estimativas, tirando uma entrevista (*7-Day Physical Activity Recall*)^{16,24,60}, todos eram questionários ($n = 18$)^{17,20,29,34,42,43,45,48,61,67,76,77,78,79,80,81,82}.

O período de referência da medida de atividade física variou de um dia^{22,23,25,26,27,30,32,35,38,44,47,56,58,59,70,72} aos últimos 12 meses⁶⁹ ou último ano^{17,42,57,73,76,81}. Os diários e as entrevistas mensuravam apenas as atividades físicas “atuais” ou “recentes”. Os questionários mensuravam tanto as atividades físicas “atuais” ou “recentes” ($n = 25$)^{19,21,22,23,34,48,50,55,61,67,68,70,77,78,79} quanto as atividades “habituais” ($n = 17$)^{17,20,24,28,42,53,57,69,71,73,76,80,81,82}. Uma semana típica/habitual ($n = 9$)^{17,20,24,53,71,80,82} e os últimos sete dias ou última semana ($n = 13$)^{21,34,43,45,48,50,55,61,67,78,79} foram os períodos mais frequentemente utilizados pelos questionários.

Reprodutibilidade

Foram identificados 41 estudos que haviam analisado a reprodutibilidade “teste-reteste” em 50 instrumentos. Trinta e dois instrumentos^{17,18,22,23,26,29,34,37,38,40,43,61,63,64,65,66,67,69,71,72,73,74,77,78,79,80,81,82} faziam parte de estudos publicados entre 2000 e 2007. A maioria foi testada em adolescentes de países da América do Norte (31/50)^{18,19,20,22,24,26,27,34,35,36,37,39,41,42,43,62,63,66,67,68,70,73,74,76,79}, sobretudo dos Estados Unidos ($n = 23$)^{18,19,22,24,26,27,34,35,36,37,41,42,43,63,66,67,68,70,73,76}, de ambos os sexos (44/50)^{17,18,19,22,23,24,27,29,34,35,36,37,38,39,40,41,42,43,61,62,63,64,65,66,67,69,70,71,72,73,74,76,77,78,79,80,81,82} e de diferentes faixas etárias (10-14 e 15-19 anos; 26/50)^{17,19,20,24,27,35,36,37,42,62,64,65,69,73,77,78,79,80,81,82}. O número de sujeitos na amostra variou de 15²³ a 4.619³⁷; a maioria tinha 50 sujeitos ou mais (34/50)^{17,18,20,22,24,29,34,35,36,37,38,39,41,43,62,64,66,67,69,71,73,74,76,77,78,79,81,82} e se tratava de amostras selecionadas por conveniência (25/50)^{17,18,22,23,24,26,27,34,36,37,40,43,62,64,66,68,69,70,71,72,73,77,79} ou por participação voluntária (8/50)^{19,38,41,42,63,67,80}.

Todas as medidas de reprodutibilidade foram estimadas mediante o procedimento “teste-

Tabela 2

Síntese dos resultados dos estudos de reprodutibilidade e validade dos instrumentos de medida da atividade física do tipo *self-report* em adolescentes (10-19 anos).

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
	Diários/Logs						
Lee & Trost ⁶⁴	3DPAR	N = 108 (M/F), Cingapura Idade: 13-16 anos	Teste-reteste, 6-8h CCI = 0,88	N = 221 (M: 105/F: 116), Cingapura Idade: 13-16 anos	Pedômetro, 4D rho = 0,40	U3D	2
McMurray et al. ²²	3DPAR	N = 320 (M: 114/F: 206), Estados Unidos Idade (média): 12,5 anos; DP = 1,1	Teste-reteste, 1D r = 0,68	N = 320 (M: 114/F: 206), Estados Unidos Idade (média): 12,5 anos; DP = 1,1	Acelerômetro, 4D r = 0,30	U3D	2
Bouchard et al. ⁶²	3DAR	N = 61 (M/F), Canadá Idade (média): 14,6 anos; DP = 2,9	Teste-reteste, 6-10D CCI = 0,91	N = 150 (M/F), Canadá Idade (média): 14,6 anos; DP = 2,9	Cicloergômetro _{PWC-150} r = 0,27	3D	2
Argiropoulou et al. ⁶¹	3DAR	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Teste-reteste, 2S CCI = 0,97	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Acelerômetro, 7D r = 0,63	3D	2
Weston et al. ³⁵	PDPAR	N = 90 (M/F), Estados Unidos Idade: 7-12 <i>grade</i> ***	Teste-reteste, 1h r = 0,98	N = 48 (M/F), Estados Unidos Idade: 7-12 <i>grade</i> ***	Pedômetro, 1D r = 0,88 Acelerômetro, 1D r = 0,77 Monitoramento da frequência cardíaca, 1D r = 0,53	UD	2
Allor & Pivarnik ²⁶	PDPAR	N = 46 (F), Estados Unidos Idade: 12 anos; DP = 0,6	Teste-reteste, 2D CCI = 0,98	N = 46 (F), Estados Unidos Idade: 12 anos; DP = 0,6	Acelerômetro, 2D r = 0,76 Monitoramento da frequência cardíaca, 2D r = 0,50	UD	2
Welky et al. ²⁵	ACTIVITYGRAM Physical Activity Self-Report			N = 28 (M/F), Estados Unidos Idade (média): 12,4 anos; DP = 0,5	Acelerômetro, 3D r = 0,50	UD	2
Pate et al. ⁵⁴	3DPAR			N = 70 (F), Estados Unidos Idade: 13-16 anos	Acelerômetro, 7D r = 0,49	3UD	2

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Bratteby et al. 46	3DAR			N = 50 (M: 25/F: 25), Suécia Idade: 15 anos	DLW, 7D não há diferença na estimativa do gasto energético total entre a DLW é o diário, $p > 0,05$ B-A, 2DP = -3,47 a 3,77 MJ/d	3D	2
Ekelund et al. 49	3DAR			N = 30 (M: 16/F: 14), Suécia Idade: 14-16 anos	Monitoramento da frequência cardíaca, 3D não há diferenças no gasto energético total e no tempo de atividade física moderada/vigorosa entre diário e o monitoramento da frequência cardíaca, $p > 0,05$ B-A, 2DP = -3,54 a 2,74 MJ/d	3D	2
Anderson et al. 44	PDPAR			N = 80 (M: 37/F: 43), Estados Unidos Idade (média): 13,4 anos; DP = 0,5	Acelerômetro, 4D $\rho = 0,42 *$	UD	2
Trost et al. 59	PDPAR			N = 37 (M: 18/F: 19), Estados Unidos Idade (média): 10,8 anos; DP = 0,1	Acelerômetro, 2D $\rho = 0,41$	UD	2
Trost et al. 58	PDPAR			N = 122 (M: 57/F: 65), Austrália Idade (média): 13,9 anos; DP = 1,2	Pedômetro, 1D $\rho = 0,34$	UD	2
Welky et al. 25	PDPAR			N = 28 (M/F), Estados Unidos Idade (média): 12,4 anos; DP = 0,4	Acelerômetro, 3D $r = 0,65$	UD	2
Rodriguez et al. 56	Self-Report Activity Diary			n = 20 (M: 13/F: 7), França Idade (média): 11,2 anos; DP = 3,8	Monitoramento da frequência cardíaca, 1D $\rho = 0,73$	UD	2

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Farias Junior et al. ⁴⁰	Adapted 3-Day Activity Record	N = 45 (M: 25/F: 20), Brasil Idade: 15-18 anos	Teste-reteste, 1D CCI = 0,84			3UD	2
	Questionários						
Sallis et al. ²⁴	GSPA	N = 102 (M: 50/F: 52), Estados Unidos Idade: 10-16 anos	Teste-reteste, 2S CCI = 0,81	N = 102 (M: 50/F: 52), Estados Unidos Idade: 10-16 anos	7DPAR, 7D r = 0,39	ST	2
Koo & Rohan ²⁰	GSPA	N = 84 (F), Canadá Idade: 7-15 anos	Teste-reteste, 11M rho = 0,48	N = 479 (F), Canadá Idade: 7-15 anos	Questionário rho = 0,33	ST	2
Gao et al. ¹⁸	GSPA	N = 250 (M: 103/F: 147), Estados Unidos Idade: 6-8 grade ***	Teste-reteste, 1S r = 0,56	N = 114 (M: 46/F: 68), Estados Unidos Idade: 6-8 grade ***	Acelerômetro, 7D r = 0,19	ST	2
Eisenmann et al. ⁶³	GSPA	N = 31 (M: 17/F: 14), Estados Unidos Idade (média): 10,6 anos; DP = 0,2	Teste-reteste, 1D r = 0,62	N = 31 (M: 17/F: 14), Estados Unidos Idade (média): 10,6 anos; DP = 0,2	Acelerômetro, 1D r = 0,50	ST	2
Telford et al. ⁷¹	CLASS	N = 111 (M: 41/F: 70), Austrália Idade: 10-12 anos	Teste-reteste, 7D CCI = 0,30	N = 111 (M: 41/F: 70), Austrália Idade: 10-12 anos	Acelerômetro, 8D rho = -0,04	ST	2
Hernández et al. ⁷⁴	CAINM	N = 114 (M: 47/F: 67), México Idade: 10-14 anos	Teste-reteste, 6M r = 0,48	N = 114 (M: 47/F: 67), México Idade: 10-14 anos	Recordatório 24h, 7D r = 0,02	UM	2
Booth et al. ⁸²	APARQ	N = 226 (M: 138/F: 88), Austrália Idade: 13-15 anos	Teste-reteste, 2S CCI = 0,69	N = 2.026 (M: 1.081/F: 945), Austrália Idade: 13 e 15 anos	Shuttle run test (número de voltas) rho = 0,22	ST	1
Ching & Diez ⁶⁸	Physical Activity Questionnaire	N = 40 (F), Estados Unidos Idade: 8,5-12,7 anos	Teste-reteste, 12-16D CCI = 0,51	N = 40 (F), Estados Unidos Idade: 8,5-12,7 anos	Diário, 7D r = 0,50	2D	2
Aaron et al. ⁷⁶	Questionnaire to Assess Past Year Physical Activity in Adolescent	N = 100 (M: 47/F: 53), Estados Unidos Idade: 15-18 anos	Teste-reteste, 12M rho = 0,69	N = 100 (M: 47/F: 53), Estados Unidos Idade: 15-18 anos	Teste da milha (minutos) rho = 0,32	UA	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Welky et al. ³⁴	YMCLS	N = 192 (M: 93/F: 99), Estados Unidos Idade: 9-13 anos	Teste-reteste, 1S CCI = 0,60	N = 192 (M: 93/F: 99), Estados Unidos Idade: 9-13 anos	Acelerômetro, 7D $r = 0,24$ Log de atividade física, 7D $r = 0,46$	U7D	1
McMurray et al. ⁷⁰	Computerized Activity Recall	N = 22 (M/F), Estados Unidos Idade (média): 11,8 anos; DP = 1,0	Teste-reteste, 1-2S CCI = 0,82	N = 45 (M: 20/F: 25), Estados Unidos Idade (média): 11,8 anos; DP = 1,0	Acelerômetro, 5D $r = 0,51$	UD	2
Prochaska et al. ⁶⁶	PACE+	N = 138 (M: 48/F: 90), Estados Unidos Idade (média): 12,1 anos; DP = 0,9	Teste-reteste, < 24h CCI = 0,77	N = 138 (M: 48/F: 90), Estados Unidos Idade (média): 12,1 anos; DP = 0,9	Acelerômetro, 5D $r = 0,40$	MC	1
Nahas et al. ²⁹	PACE+	N = 122 (M: 44/F: 78), Brasil Idade: 15-24 anos	Teste-reteste, 1S CCI = 0,93	N = 65 (M: 26/F: 39), Brasil Idade: 15-24 anos	Pedômetro, 7D $\rho = 0,23$ Recordatório, 7D $\rho = 0,41$	MC	1
Ridley et al. ²³	CDPAQ	N = 15 (M: 7/F: 8), Austrália Idade (média): 12,1 anos; DP = 0,5	Teste-reteste, < 24h CCI = 0,94	N = 30 (M: 15/F: 15), Austrália Idade (média): 11,96 anos; DP = 0,5	Acelerômetro, 1D $r = 0,45$ Monitoramento da frequência cardíaca, 1D $r = 0,66$	UD	2
Argiropoulou et al. ⁶¹	PALQ	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Teste-reteste, 2S CCI = 0,52	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Acelerômetro, 7D $r = 0,53$	U7D	1
Philippaerts et al. ⁸⁰	Flemish Physical Activity Questionnaire	N = 33 (M: 10/F: 23), Bélgica Idade: 12-18 anos	Teste-reteste, 9D CCI = 0,68	N = 33 (M: 10/F: 23), Bélgica Idade: 12-18 anos	Acelerômetro, 7D $r = 0,45$	ST	1
Barbosa et al. ⁸¹	QAPACE	N = 121 (M: 56/F: 65), Colômbia Idade: 8-16 anos	Teste-reteste, 90D CCI = 0,96	N = 36 (M: 18/F: 18), Colômbia Idade: 8-16 anos	VO ₂ max $r = 0,63$	UA	1
Wong et al. ⁷⁹	SHAPES	N = 1.636 (M: 729/F: 904), Canadá Idade: 9-12 grade ***	Teste-reteste, 1S $\kappa = 0,37$	N = 67 (M: 34/F: 33), Canadá Idade: 9-12 grade ***	Acelerômetro, 7D $\rho = 0,44$	U7D	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Guedes et al. ⁷⁷	IPAQ	N = 161 (M: 69/F: 92), Brasil Idade: 12-18 anos	Teste-reteste, 2S rho = 0,59	N = 161 (M: 69/F: 92), Brasil Idade: 12-18 anos	3DAR, 4D rho = 0,26	US	1
Florindo et al. ¹⁷	ST	N = 94 (M: 30/F: 64), Brasil Idade: 11-16 anos	Teste-reteste, 15D CCI = 0,61	N = 94 (M: 30/F: 64), Brasil Idade: 11-16 anos	<i>Shuttle run test</i> rho = 0,15	ST	1
	UA	N = 94 (M: 30/F: 64), Brasil Idade: 11-16 anos	Teste-reteste, 15D CCI = 0,68	N = 94 (M: 30/F: 64), Brasil Idade: 11-16 anos	<i>Shuttle run test</i> rho = 0,23	UA	1
Booth et al. ⁷⁸	WHO HBSC	N = 226 (M: 138/F: 88), Austrália Idade: 13,7-15,7 anos	Teste-reteste, 2S kappa = 0,38	N = 2.026 (M: 1.082/F: 944), Austrália Idade: 13,7-15,7 anos	<i>Shuttle run test</i> , maior número de voltas nos adolescentes fisicamente mais ativos	U7D	1
Guedes et al. ⁶⁹	<i>Baecke Questionnaire of Habitual Physical Activity</i>	N = 161 (M: 69/F: 92), Brasil Idade: 12-18 anos	Teste-reteste, 2S CCI = 0,73	N = 161 (M: 69/F: 92), Brasil Idade: 12-18 anos	Diário, 4D k = 0,44	U12M	2
Koo & Rohan ²⁰	PS	N = 84 (F), Canadá Idade: 7-15 anos	Teste-reteste, 11M rho = 0,44	N = 479 (F), Canadá Idade: 7-15 anos	Questionário rho = 0,29	ST	2
	SS	N = 84 (F), Canadá Idade: 7-15 anos	Teste-reteste, 11M rho = 0,59	N = 479 (F), Canadá Idade: 7-15 anos	Questionário rho = 0,16	ST	2
	SAS	N = 84 (F), Canadá Idade: 7-15 anos	Teste-reteste, 11M rho = 0,53	N = 479 (F), Canadá Idade: 7-15 anos	Questionário rho = 0,26	ST	1
Janz et al. ¹⁹	3DSR	N = 30 (M: 15/F: 15), Estados Unidos Idade: 7-15 anos	Teste-reteste, 1M CCI = 0,30	N = 30 (M: 15/F: 15), Estados Unidos Idade: 7-15 anos	Acelerômetro, 6D r = 0,48	3UD	2
	3DAR	N = 30 (M: 15/F: 15), Estados Unidos Idade: 7-15 anos	Teste-reteste, 1M CCI = 0,54	N = 30 (M: 15/F: 15), Estados Unidos Idade: 7-15 anos	Acelerômetro, 6D r = 0,39	3UD	2
Troped et al. ⁶⁷	<i>Youth Risk Behavior Survey Questionnaire – 2003</i>	N = 128 (M/F), Estados Unidos Idade (média): 12,6 anos; DP = 0,6	Teste-reteste, 1-2S CCI = 0,49	N = 125 (M: 60/F: 65), Estados Unidos Idade (média): 12,6 anos; DP = 0,6	Acelerômetro, 7D SE = 0,20 e ES = 0,82	U7D	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Argiropoulou ⁶¹	<i>Four by One-Day Recall Physical Activity Questionnaire</i>	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Teste-reteste, 2S CCI = 0,70	N = 40 (M: 23/F: 17), Grécia Idade: 13-14 anos	Acelerômetro, 7D $r = 0,62$	3UD	2
Treuth ⁷³	<i>Fels Physical Activity Questionnaire</i>	N = 229 (M: 99/F: 130), Estados Unidos Idade: 7-19 anos	Teste-reteste, 6D CCI = 0,66	N = 229 (M: 99/F: 130), Estados Unidos Idade: 7-19 anos	Acelerômetro, 6D $\rho = 0,21$	UA	2
Ridley et al. ⁷²	MARCA	N = 32 (M: 18/F: 14), Austrália Idade: 9-13,5 anos	Teste-reteste, 1D CCI = 0,94	N = 66 (M: 33/F: 33), Austrália Idade: 9-15 anos	Acelerômetro, 1D $\rho = 0,35$	UD	2
Garcia et al. ²⁷	CAAL	N = 25 (M/F), Estados Unidos Idade: 11-15 anos	Teste-reteste, 1D CCI = 0,95	N = 94 (M/F), Estados Unidos Idade: 11-15 anos	Acelerômetro, 5D $r = 0,39$ <i>Step test</i> $r = 0,14$	UD	2
McMurray et al. ²²	SAPAC	N = 320 (M: 206/F: 114), Estados Unidos Idade (média): 12,5 anos; DP = 1,1	Teste-reteste, 1D $r = 0,67$	N = 320 (M: 206/F: 114), Estados Unidos Idade (média): 12,5 anos; DP = 1,1	Acelerômetro, 4D $r = 0,24$	UD	1
Mota et al. ⁶⁵	Weekly activity checklist	N = 30 (M/F), Portugal Idade: 8-16 anos	Teste-reteste, 1S CCI = 0,70	N = 109 (M: 42/F: 67), Portugal Idade: 8-16 anos	Acelerômetro, 3D $r = 0,30$	US	2
Scerpella et al. ³¹	GSPA			N = 61 (F), Estados Unidos Idade (média): 10,2 anos; DP = 1,1	<i>Checklist</i> , 6D $\rho = 0,25$ Acelerômetro, 3D $\rho = 0,10$	ST	2
Ridder et al. ⁵⁵	<i>Weight Bearing Activity Questionnaire</i>			N = 72 (M: 35/F: 37), Holanda Idade: 8-14 anos	Acelerômetro, 4-5D $\rho = 0,44$	U7D	2
Verheul et al. ³³	<i>Weight Bearing Activity Questionnaire</i>			N = 26 (F), Holanda Idade: 11-15 anos	Acelerômetro, 3D, $r = 0,12$ 3DAR, 3D $r = 0,57$	U7D	2
Ekelund et al. ⁴⁸	SAPAQ			n = 49 (M: 18/F: 31), Suécia Idade (média): 16,8 anos; DP = 0,4	Acelerômetro, 7D $r = 0,51$	U7D	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Kowalski ⁵⁰	PAQ-A			N = 85 (M: 41/F: 44), Canadá Idade: 13-20 anos	Acelerômetro, 7D r = 0,33 7DPAR r = 0,51 GSPA, 7D r = 0,57	U7D	2
Kowalski et al. ²¹	PAQ-C			N = 97 (M: 41/F: 56), Canadá Idade: 9-14 anos	7-day recall, 7D r = 0,43 Acelerômetro, 7D r = 0,39 GSPA, 7D r = 0,41 Step test r = 0,28	U7D	2
	PAQ-C			N = 89 (M: 38/F: 51), Canadá Idade: 8-13 anos	Tempo de atividade física moderada/vigorosa, 7D r = 0,47	U7D	2
Moore et al. ⁵²	PAQ-C			N = 991 (M: 494/F: 490), Estados Unidos Idade: 8-14 anos	Step test r = 0,08	U7D	2
Crocker et al. ¹⁶	PAQ-C			N = 61 (M: 27/F: 34), Canadá Idade (média): 11,2 anos; DP = 0,6	Acelerômetro – R3D, 7D r = 0,13 7DPAR, 7D r = 0,39	U7D	2
Gao et al. ¹⁸	Youth Risk Behavior Survey Questionnaire – 1999			N = 114 (M: 46/F: 68), Estados Unidos Idade: 6-8 grade ***	Acelerômetro, 7D r = 0,10	U7D	1
Slinde et al. ⁵⁷	Minnesota Leisure Time Physical Activity Questionnaire			N = 35 (M: 18/F: 17), Suécia Idade: 15-17 anos	DLW, 14D rho = 0,49	UA	2
Arvidsson et al. ⁴⁵	Physical Activity Questionnaire for Adolescent			N = 33 (M: 17/F: 16), Suécia Idade: 15 anos	DLW, 14D r = 0,62	U7D	1
Sallis et al. ³⁰	SAPAC			N = 125 (M: 55/F: 70), Estados Unidos Idade (média): 10,9 anos; DP = 0,5	Monitoramento da frequência cardíaca, 1D r = 0,31 Acelerômetro, 1D r = 0,54	UD	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Kemper et al. ²⁸	PAQ			N = 200 (M/F), Holanda Idade: 13-16 anos	Pedômetro, 4D $r = 0,20$ Monitoramento da frequência cardíaca, 2D $r = 0,17$	3UM	2
Murphy et al. ⁵³	Three Activity Posteres			N = 92 (M/F), Estados Unidos Idade: 10-18 anos	VO ₂ max $\rho = 0,21$	ST	2
Crocker ³⁹	PAQ-C	N = 84 (M/F), Canadá Idade: 9-14 anos	Teste-reteste, 1S CCI = 0,78			U7D	2
Breener et al. ³⁷	Youth Risk Behavior Survey Questionnaire – 1999	N = 4.619 (M: 2.152/F: 2.457), Estados Unidos Idade: 9-12 grade ***	Teste-reteste, 10-22D $\kappa = 0,61$			U7D	1
Zulling et al. ⁴³	Middle School Youth Risk Behavior Questionnaire – 2005	N = 232 (M: 105/F: 127), Estados Unidos Idade: 7-8 grade ***	Teste-reteste, 2S $\kappa = 0,76$			U7D	1
Brener et al. ³⁶	Youth Risk Behavior Survey Questionnaire – 1992	N = 1.679 (M: 774/F: 905), Estados Unidos Idade: 7-12 grade ***	Teste-reteste, 2S $\kappa = 0,75$			U7D	1
Gilmer et al. ⁴¹	Weekly Activity Checklist	N = 103 (M/F), Estados Unidos Idade (média): 12,3 anos; DP = 1,0	Teste-reteste, 2S $r = 0,70$			US	2
Brown & Holland ³⁸	Modified Version Self-Assessed Physical Activity Checklist	N = 103 (M: 52/F: 51), Austrália Idade (média): 11,7 anos; DP = 0,5	Teste-reteste, 5D CCI = 0,20			UD	1
Kriska et al. ⁴²	MAQ Entrevistas estruturadas	N = 23 (M: 10/F: 13), Estados Unidos Idade: 10-59 anos	Teste-reteste, 1-3S $\rho = 0,37$			UA	1

(continua)

Tabela 2 (continuação)

Referências	Tipo e nome original do instrumento	Reprodutibilidade "teste-reteste"		Validade		Período *	Capacidade **
		Características da amostra	Métodos e resultados	Características da amostra	Critério de referência e resultados		
Sallis et al. ²⁴	7DPAR	N = 102 (M: 50/F: 52), Estados Unidos Idade (média): 11,2 anos; DP = 0,6 e 16,4; DP = 0,7	Teste-reteste, 6D CCI = 0,77	N = 102 (M: 50/F: 52), Estados Unidos Idade (média): 11,2 anos; DP = 0,6 e 16,4; DP = 0,7	Monitoramento da frequência cardíaca, 1D r = 0,49	U7D	1
Wallace et al. ⁶⁰	7DPAR			N = 11 (M), Estados Unidos Idade (média): 12,5 anos; DP = 0,7	Observação direta do comportamento, US Não havia diferenças significativas no gasto energético total entre os dois métodos	U7D	1
Crocker et al. ¹⁶	7DPAR			N = 61 (M: 27/F: 34), Canadá Idade (média): 11,2 anos; DP = 0,6	Acelerômetro – R3D, 7D r = -0,13 PAQ-C, U7D r = 0,39	U7D	1
Simons-Morton et al. ³²	Physical Activity Record			N = 48 (M: 20/F: 28), Estados Unidos Idade: 7-11 anos	Acelerômetro, 1D r = 0,54 Monitoramento da frequência cardíaca, 1D r = 0,62	UD	2
Larsson et al. ⁵¹	PAI			N = 32 (M: 17/F: 15), Suécia Idade: 16-20 anos	DLW, 14D PAI subestima o gasto energético total nos mais ativos e superestima nos menos ativos; 1,4 ± 2,6 MJ/d (IC95%: 2,4-0,5 MJ/d).	2US	2

(continua)

Validade

Cinquenta e oito estudos haviam analisado a validade de 71 instrumentos. Quarenta e seis^{16,17,18,22,23,25,26,28,29,31,34,44,45,47,48,51,52,54,55,56,57,58,61,63,64,65,66,67,69,71,72,73,74,77,78,79,80,81,82} eram referentes a estudos publicados entre 2000 e 2007; a maioria desses instrumentos (43/71)^{16,18,19,20,21,22,24,25,26,27,30,31,32,34,35,44,47,50,52,53,54,59,60,62,63,66,67,68,70,73,74,76,79} foi testada em adolescentes de poucos países da América do Norte, especialmente Estados Unidos (n = 31)^{18,19,22,24,25,26,27,30,31,32,34,35,44,47,52,53,54,59,60,63,66,67,68,70,73,76}, de ambos os sexos (61/71), mas não de diferentes faixas etárias (10-14 e 15-19 anos, 41/71). O tamanho da amostra variou de 11⁶⁰ a 2.026 sujeitos^{78,82}. A maioria tinha 50 sujeitos ou mais (45/71)^{16,17,18,20,21,22,24,27,28,29,30,31,34,44,46,50,52,53,54,55,58,62,64,65,66,67,69,71,72,73,74,76,77,78,79,82} e se tratava de amostras por conveniência (42/71)^{17,18,21,22,23,24,26,27,30,31,32,33,34,44,46,48,49,50,51,52,53,54,57,58,59,60,62,64,66,67,68,69,70,71,72,73,77,79} ou por participação voluntária (12/71)^{16,19,25,28,45,56,63,65,80}.

Quinze instrumentos^{16,21,23,26,27,28,29,30,31,32,33,34,35,50} haviam sido testados contra mais de um critério de referência – “padrão-ouro”, o que resultou em 90 medidas de validade. O coeficiente de correlação de Pearson foi a análise mais utilizada para estimar a validade dos instrumentos (58/90)^{16,18,19,21,22,23,24,25,26,27,28,30,32,33,34,35,45,47,48,50,52,54,61,62,63,65,66,68,70,74,80,81}, seguido pela correlação de Spearman (25/90)^{9,17,20,29,31,44,53,55,56,57,58,59,64,71,72,73,76,77,79,82}. O diagrama de dispersão de Bland-Altman^{46,49,51}, o índice kappa⁶⁹, o teste t para amostras independentes^{46,51,60} e medidas de sensibilidade e especificidade⁶⁷ também foram utilizados.

Cinquenta e nove medidas de validade^{16,18,19,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,44,45,47,48,50,51,54,55,56,57,58,59,60,61,63,64,65,66,67,70,71,72,73,79,80} tiveram como “padrão-ouro” ao menos uma medida objetiva da atividade física. Dentre elas, a medida de atividade física por meio de acelerômetros foi o critério mais utilizado (40/90)^{16,18,19,21,22,23,24,25,26,27,31,32,33,34,35,44,47,48,50,54,55,59,61,63,65,66,67,70,71,72,73,79,80}. As medidas subjetivas foram relativamente pouco utilizadas como critério de referência (20/90; 22%)^{16,20,21,24,29,33,34,50,68,69,74,77}, sobretudo, como o critério principal (10/90)^{20,21,24,68,69,74,77}. Medidas de aptidão cardiorrespiratória também foram utilizadas como “padrão-ouro” para avaliar os níveis de validade dos instrumentos (testes de campo e de laboratório – medidas diretas e indiretas) (10/90)^{17,21,27,52,53,62,76,78,81,82}. A combinação de medidas de atividade física também foi adotada como “padrão-ouro”: de medidas objetivas^{23,26,28,30,32,35}, objetivas e subjetivas

^{16,21,29,31,33,34,50}, e destas com de aptidão cardiorrespiratória^{21,27}.

As medidas de validade variaram de -0,13¹⁶ a 0,88³⁵; a maioria (64/84; 76%)^{16,17,18,19,20,21,22,24,25,26,27,28,29,30,31,33,34,44,47,50,52,53,54,55,57,58,59,62,63,64,65,66,68,69,71,72,73,74,76,77,79,80,82} dos coeficientes de correlação foi igual ou inferior a 0,50; 16^{21,23,25,30,32,33,35,45,48,50,61,70,81} tinham valores entre 0,51-0,69; quatro alcançaram valores iguais ou superiores a 0,70^{26,35,56}.

Parece haver uma relação entre a magnitude dos coeficientes de correlação para a validade e o tipo de critério de referência utilizado (Figura 2). Quase todas as correlações foram $\leq 0,50$ quando se compararam os instrumentos às medidas subjetivas da atividade física (17/20; 85%)^{16,20,21,24,29,31,34,68,69,74,77} ou de aptidão cardiorrespiratória (9/10; 90%)^{17,21,27,52,53,62,76,78,82}. Nas medidas objetivas de atividade física, verificou-se uma menor frequência de correlações de magnitude muito baixa ($< 0,30$) e baixa ($\leq 0,50$) (38/54; 70,1%).

Ao considerar apenas os estudos que utilizaram medidas objetivas mais robustas (DWL, acelerômetros, monitores de frequência cardíaca) como “padrão-ouro”, verificou-se que a maioria dos coeficientes de correlação – medida de validade – (32/44; 72,7%) foi igual ou inferior a 0,50 (diários: 50%; questionários: 80%; entrevistas: 75%); 20,5% (9/44) entre 0,51-0,69 e 6,8% (3/44) com valores iguais ou superiores a 0,70.

Dentre as medidas de validade que foram comparadas segundo o sexo, idade e nível de intensidade das atividades físicas, a maioria evidenciou maiores níveis de validade nas medidas dos adolescentes do sexo masculino (15/22)^{17,22,28,29,33,44,45,55,63,66,69,70,77,81}, nos de maior idade (15-19 anos, 7/10)^{24,32,34,65,69,77,81} e nas atividades físicas vigorosas (15/18)^{18,22,24,34,54,58,59,64,67,73,74,76,77,80}.

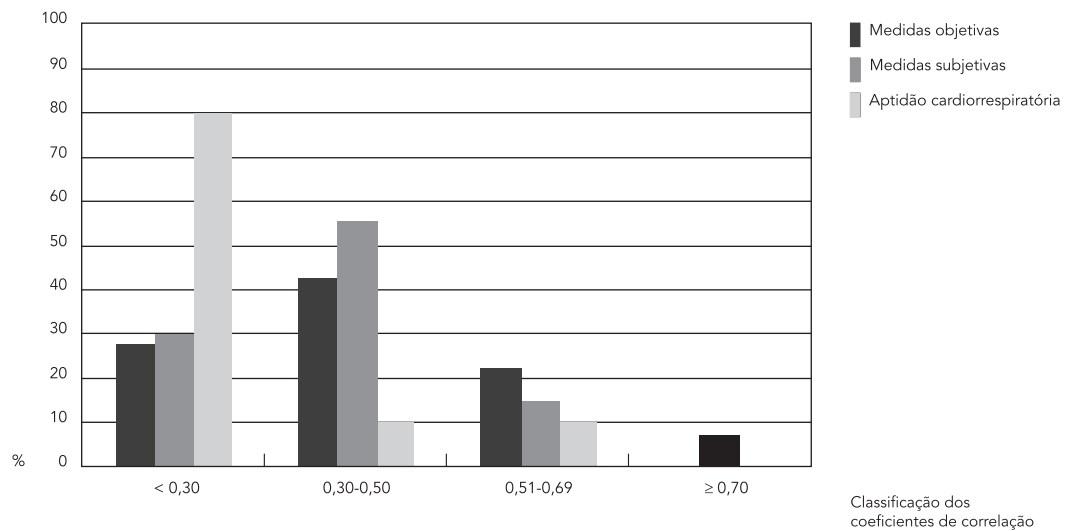
Discussão

Há uma grande diversidade de instrumentos de medida da atividade física do tipo *self-report* que foram testados (reprodutibilidade e validade) em adolescentes (10-19 anos), sobretudo questionários. Tais instrumentos mensuram de diferentes formas a atividade física, por diferentes períodos de tempo. Em geral, demonstraram maior reprodutibilidade “teste-reteste” do que validade. Maiores coeficientes de reprodutibilidade e validade foram relativamente mais frequentes nos diários do que nos questionários e entrevistas.

Trata-se da primeira revisão sistemática a analisar, de forma simultânea, diferentes carac-

Figura 2

Classificação dos coeficientes de correlação: medidas de validade.



terísticas dos instrumentos *self-report* de medida da atividade física em adolescentes (além da reprodutibilidade e da validade) e os aspectos metodológicos desses estudos. Isto poderá ajudar pesquisadores e profissionais que pretendam mensurar a atividade física em adolescentes a responder uma questão comum: “Qual o instrumento devo utilizar?”.

Este estudo apresenta algumas limitações que precisam ser levadas em conta. Uma das limitações foi não ter avaliado a qualidade metodológica dos estudos incluídos na revisão. Entretanto, destaca-se que os principais aspectos metodológicos que são avaliados (seleção da amostra, análise dos dados, qualidade das medidas utilizadas) fizeram parte da análise desta revisão. Além disso, as implicações das possíveis limitações dos estudos sobre os níveis de reprodutibilidade e validade dos instrumentos foram analisadas e discutidas.

Outra limitação foi não ter incluído os estudos que apresentavam os coeficientes de reprodutibilidade e/ou validade como parte dos métodos em estudos com outros objetivos. No entanto, acredita-se que os resultados não seriam diferentes, caso estes estudos tivessem sido incluídos. Deve-se considerar, também, a possibilidade de algum estudo que atendia aos critérios de inclusão não ter sido identificado pelos métodos de busca.

Mensurar a atividade física dos adolescentes, apesar de ser uma tarefa desafiadora, tem despertado o interesse de diversos pesquisadores em vários países, como indicado pelo número de instrumentos desenvolvidos e testados ($n = 52$), e de estudos publicados sobre as propriedades psicométricas desses instrumentos ($\approx 65\%$ entre 2000 e 2007).

Contudo, na maioria das vezes, semelhantes estudos parecem que mais contribuíram para adicionar novos instrumentos de medida da atividade física do que sistematicamente exploraram princípios de medida da atividade física em adolescentes. Por exemplo, apenas 11 instrumentos, dos 52 identificados nesta revisão, foram testados mais de uma vez. Além disso, a maioria dos estudos desenvolvidos ficou restrita a poucos países da América do Norte, não incluiu adolescentes de diferentes faixas etárias e utilizou sujeitos selecionados por conveniência ou participação voluntária.

Em relação às implicações da adoção desse tipo de seleção da amostra, não se pode descartar a hipótese de que os sujeitos sejam participantes mais motivados, podendo artificialmente inflar as medidas de reprodutibilidade e/ou de validade^{83,84}. Outro aspecto que merece ser destacado é a possibilidade de ocorrência de “viés cultural”⁸⁴, caso esses instrumentos sejam aplicados diretamente (sem adaptação e testagem

prévias) em amostras com características distintas das quais os mesmos foram desenvolvidos e avaliados. Dentre os 52 instrumentos identificados, apenas cinco^{17,29,40,69,77} foram testados em adolescentes brasileiros.

A proliferação de instrumentos do tipo *self-report* em adolescentes pode indicar três aspectos: a diversidade de interesses de pesquisa envolvendo a prática de atividade física; a relação “dose-resposta” da atividade física difere segundo o desfecho em estudo, exigindo medidas específicas; falta de consenso sobre quais instrumentos utilizar em determinadas situações.

A vantagem de se dispor de múltiplos instrumentos é a possibilidade de responder a diferentes objetivos em estudos com atividade física. Por sua vez, isso pode dificultar ou impossibilitar a comparação e interpretação de resultados dos estudos, pois há diferenças, às vezes acentuadas, na forma como os instrumentos operacionalizam e expressam o nível de atividade física dos adolescentes^{8,12,83}.

Reprodutibilidade

O procedimento de medidas repetidas foi utilizado para estimar todas as medidas de reprodutibilidade, conforme descrito em outras revisões^{8,9,14,15}. Observou-se, porém, ampla variação no intervalo entre as réplicas de aplicação dos instrumentos, utilizando-se desde poucas horas até vários meses. Recomenda-se que o intervalo transcorrido entre as aplicações seja curto, de um a três dias, no máximo uma semana, e que haja sobreposição no período de recordação das atividades físicas praticadas nas duas aplicações^{15,83,84}.

Desse modo, parece que vários estudos não examinaram a reprodutibilidade “teste-reteste” de forma adequada, pois utilizaram intervalos prolongados (mais de uma semana, $n = 27$)^{17,19,20,24,36,37,41,42,43,61,67,68,69,70,74,76,77,78,81,82} ou diferentes períodos de recordação nas duas aplicações ($n = 23$)^{19,24,29,34,36,37,38,39,41,43,61,62,65,67,68,70,74,77,78,79}. Em função disso, muitos estudos estão tratando mudanças naturais na atividade física, ocorridas entre as réplicas de aplicação, como fonte de erro na estimativa da reprodutibilidade.

A maioria das medidas de reprodutibilidade “teste-reteste” foi estimada pelo CCI. Essa análise é considerada a mais apropriada para tal finalidade (dados contínuos e distribuição normal)^{15,83,85}. A correlação de Spearman foi utilizada para dados que não apresentaram uma distribuição normal e o índice kappa para dados categóricos (exemplo, suficientemente ativos *vs.* insuficientemente ativos).

Um dado preocupante em relação à análise estatística foi que 14%^{18,22,35,41,63,74} das medidas de reprodutibilidade foram estimadas pelo coeficiente de correlação de Pearson. Essa análise não deve ser utilizada para estimar reprodutibilidade. Primeiro, por se tratar de uma estatística bivariada, enquanto a situação em questão requer uma análise univariada^{15,83,85}; segundo, essa análise mede a magnitude da relação linear entre duas variáveis, sendo pouco sensível a pequenas variações que, normalmente, ocorrem no nível de atividade física entre as réplicas de aplicação do instrumento^{15,83,85}.

Houve uma grande variabilidade na medida de reprodutibilidade. Constatou-se que alguns instrumentos apresentaram apenas baixos níveis de reprodutibilidade ($\leq 0,50$)^{19,20,38,42,67,71,74,78,79} e outros níveis elevados ($> 0,95$)^{26,27,35,61,81}. Vinte e dois instrumentos^{23,24,26,27,29,35,36,39,40,41,43,61,62,64,65,66,70,72,77,81} demonstraram níveis de reprodutibilidade que podem ser considerados como satisfatórios ($\geq 0,70$)^{84,89} e 11^{17,22,34,37,63,73,76,80,82} tiveram níveis moderados (0,60-0,69)^{83,85}.

Não há diferenças significantes na classificação dos níveis de reprodutibilidade entre a análise com e sem a inclusão dos coeficientes determinados por meio da correlação de Pearson ($< 0,70$: 28/50, 56% *vs.* $< 0,70$: 23/43, 53,5%). O mesmo foi observado para os coeficientes de maior magnitude ($> 0,90$: 9/50, 18% *vs.* 8/43, 18,6%). Também não houve mudanças na amplitude dos coeficientes, permanecendo entre 0,20-0,98.

Baixos níveis de reprodutibilidade observados na maioria dos instrumentos pode ser uma indicação de que esses instrumentos apresentam problemas de clareza e objetividade. Outros fatores como a falta de padronização, mudanças na atividade física entre as réplicas de aplicações dos instrumentos também precisam ser considerados.

Maior reprodutibilidade “teste-reteste” nos diários pode ser atribuída à estrutura dos instrumentos (duração mensurada em blocos: 15min, 30min; atividades informadas em códigos/grupos) e ao fato de eles mensurarem as atividades físicas praticadas mais recentemente (exemplo, último dia, os últimos dois ou três dias)^{4,8,12}.

Poucos estudos ($< 1/3$) compararam os níveis de reprodutibilidade por sexo, idade e intensidade da atividade física. Dentre os que compararam, a maioria evidenciou maiores níveis de reprodutibilidade nas informações fornecidas pelos rapazes, pelos adolescentes mais velhos (15-19 anos) e nas atividades físicas de intensidade vigorosa. Possivelmente, isso se deve à maior participação dos rapazes em atividades físicas vigorosas, comparados com as moças, que tendem a ter maior participação em atividades

físicas moderadas^{4,8,12,75}. Normalmente, as atividades vigorosas são estruturadas, planejadas e realizadas de forma repetitiva, sendo mais fáceis de serem recordadas³.

Em relação às possíveis influências da idade, acredita-se que o padrão de atividades físicas e as características psicológicas e cognitivas expliquem os resultados encontrados^{3,8}. Em geral, os adolescentes mais jovens apresentam maior envolvimento em atividades físicas esporádicas e de caráter intermitente, comparados aos seus pares mais velhos^{8,12}. Estas atividades são mais difíceis de serem quantificadas e, portanto, recordadas³.

Níveis mais elevados de reprodutibilidade foram identificados nos instrumentos que mensuravam atividades físicas “recentes” comparados àqueles que mensuravam atividades habituais. Esses resultados reforçam indicações da literatura sobre a relação inversa entre precisão e tempo de recordação das atividades físicas^{3,4,8,12,15,83,84}. A precisão com que as atividades físicas são recordadas diminui conforme aumenta o período de recordação.

Validade

A ausência de um “padrão-ouro”, que seja adequado e consensual, contra o qual os instrumentos de medida da atividade física do tipo *self-report* – medidas subjetivas – sejam comparados ainda representa uma das principais lacunas nessa área^{3,4,8,12,15,83,84}. No presente estudo, foram identificados nove critérios de referência utilizados na validação: acelerômetros, pedômetros, monitores da frequência cardíaca, DLW, observação direta do comportamento, diários, questionários, entrevistas e medida da aptidão cardiorrespiratória.

Todos esses métodos apresentam algum tipo de limitação bem como mensuram de diferentes formas o nível de atividade física^{4,5,8,12,15}, produzindo diferentes resultados em termos de validade. Tal hipótese pode ser sustentada pelos achados desta revisão, que observou correlações de diferentes magnitudes para o mesmo instrumento (na mesma amostra), entre as medidas objetivas^{23,26,28,30,32,35}, subjetivas^{21,50}, objetivas e subjetivas^{16,21,29,31,33,34,50} e entre estas e as de aptidão cardiorrespiratória^{21,27}.

Por exemplo, os acelerômetros (“padrão-ouro” mais utilizado) apresentam algumas particularidades e limitações: (1) diferentes equações têm sido utilizadas para determinar a intensidade das atividades físicas, produzindo resultados divergentes^{44,86}; (2) a validade desses equipamentos e as suas equações foram desenvolvidas em laboratório, podendo não ter o mesmo de-

sempenho em condições de “vida real”^{8,12}; (3) mensuram com limitações algumas atividades e não podem ser utilizados em outras^{8,12,86}; (4) mensuram apenas atividades físicas “recentes” ou “atuais”.

O coeficiente de correlação de Pearson e de Spearman foram as análises estatísticas mais utilizadas para determinar validade dos instrumentos. Essas análises são consideradas indicadores limitados de validade, uma vez que permitem avaliar, essencialmente, a magnitude da força da relação linear entre duas variáveis, mas não a concordância entre elas^{83,87}. Elas também não permitem avaliar a ocorrência de erro sistemático ou aleatório⁸⁷. Uma alternativa complementar é utilizar, quando possível (considerar pressupostos), medidas de sensibilidade/especificidade, o índice kappa e o método de Bland-Altman.

Uma grande variação foi observada nos coeficientes de validade, -0,13 a 0,88. Essa variabilidade é um indicativo de que muitos instrumentos e critérios de referência têm sido utilizados⁸. A maioria dos instrumentos apresentou coeficientes de correlação (medida de validade) de baixa magnitude ($\leq 0,50$)^{15,84}. Dentre os instrumentos analisados, apenas três apresentaram correlações $\geq 0,70$, todos eram diários: *Previous Day Physical Activity Diary* (PDPAR)^{26,35} e *Self-Report Activity Diary*⁵⁶. Todos comparados com uma medida objetiva.

Os maiores coeficientes de correlação encontrados nos questionários variaram de 0,62 a 0,66. Cinco questionários alcançaram tais níveis de correlação: *Computer Delivered Physical Activity Questionnaire* (CDPAQ, $r = 0,66$); *Four by One-Day Recall Physical Activity Questionnaire* ($r = 0,62$); *Physical Activity Questionnaire* ($r = 0,63$); *Quantification de l'Activite Physique en Altitude Chez le Enfants* (QAPACE, $r = 0,63$). Dentre as entrevistas analisadas, apenas uma apresentou correlação superior a 0,60 (*Physical Activity Record*, $r = 0,62$). Medidas objetivas de atividade física serviram de critério de referência para avaliar todos os instrumentos, exceto o QAPACE (medida de VO_2 máx por ergoespirometria).

Correlações de maior magnitude foram relativamente mais frequentes quando o critério de referência foi uma medida objetiva de atividade física, comparados às medidas subjetivas e de aptidão cardiorrespiratória. Esses resultados precisam ser interpretados com bastante cautela, dado que se trata, na maioria das vezes, de diferentes instrumentos, avaliados em amostras com diferentes características (por exemplo, faixa etária, número de sujeitos).

Particularmente em relação aos resultados dos estudos que utilizaram uma medida de aptidão cardiorrespiratória, os baixos coeficientes

de correlação podem ser explicados pela grande carga genética desse indicador e por se tratar de uma medida de aptidão física e não de atividade física^{4,8,9}.

Medidas objetivas de atividade física têm sido recomendadas para servir de critério de referência na validação das medidas subjetivas^{4,5,8}. Pressupõe-se que esses métodos produzam estimativas mais robustas do nível de atividade física dos adolescentes, pois, não são influenciadas por vieses de “memória” e de “estimação”^{4,8,86}.

Nesse sentido, realizou-se uma análise das evidências de validade dos instrumentos, levando em consideração apenas os estudos que tiveram como “padrão-ouro” uma medida objetiva “robusta” (acelerômetros, DLW, monitores de frequência cardíaca). Desse modo, 44 instrumentos foram analisados. A maioria (72,7%) das correlações foi igual ou inferior a 0,50 (diários: 50%; questionários: 80%; entrevistas: 75%). Considerando as correlações de magnitude moderada (0,51-0,69) e moderada a elevada ($\geq 0,70$), 18% (n = 9) e 6,8% (n = 3), esses instrumentos apresentaram, respectivamente, tais coeficientes.

Estes resultados evidenciaram que mesmo utilizando apenas critérios de referências que podem ser considerados mais “robustos”, a maioria dos instrumentos analisados nesta revisão apresentou baixos níveis de validade, sobretudo os questionários e as entrevistas. Melhores evidências de validade foram produzidas quando o “padrão-ouro” adotado foi uma medida objetiva de atividade física. Todavia, independentemente do “padrão-ouro”, a maioria dos instrumentos analisados tem níveis de validade questionáveis.

Os baixos coeficientes de correlação que foram identificados na grande maioria dos instrumentos podem ser explicados de três formas: primeira, os instrumentos mensuram com limitações o nível de atividade física, pois dependem da capacidade dos adolescentes de recordar, interpretar e quantificar as atividades físicas praticadas – medidas subjetivas; segundo, os métodos de medida da atividade física apresentam limitações. Alguns instrumentos foram testados contra outras medidas subjetivas, cujos níveis de validade são desconhecidos. Isso tem sido um dos maiores problemas nessa área. Problemas de precisão no método de referência tendem a produzir subestimação da validade do instrumento que está sendo testado ou avaliado⁸⁸. Terceiro, deve-se considerar que há, em várias ocasiões, uma “incompatibilidade” temporal (atividades “recentes” *vs.* típicas/habituais) e na operacionalização da medida de atividade física entre a medida do “padrão-ouro” e a do instrumento que está sendo testado. Os instrumentos mensuram

de diferentes formas diferentes aspectos da atividade física.

Os diários apresentaram níveis mais elevados de validade do que os questionários e entrevistas, conforme descrito na literatura^{4,8,9,12}. Esses instrumentos foram os únicos que apresentaram correlações superiores a 0,70. Isso se deve, provavelmente, à forma como eles mensuram a atividade física (as mais recentes – minimiza “viés de memória”; as atividades são referidas em categorias/grupos; a duração é mensurada em blocos: 15min, 30min) e a utilização de uma medida objetiva como critério de referência (acelerômetros).

A maioria dos estudos relatou maior validade nas medidas fornecidas pelos rapazes, os adolescentes mais velhos (15-19 anos) e nas atividades físicas vigorosas, comparados às moças, aos adolescentes mais jovens (10-14 anos) e às atividades moderadas. Resultados similares foram descritos previamente¹⁵. As possíveis explicações para esses achados são semelhantes às apresentadas para a medida de reprodutibilidade.

Foram identificados 52 instrumentos diferentes, a maioria deles questionários. Esses instrumentos mensuram de diferentes formas as dimensões e domínios da atividade física, por períodos de tempo que variam de um dia até os últimos 12 meses. Alguns instrumentos produzem uma medida geral de atividade física, outros fornecem informações detalhadas sobre as atividades físicas praticadas em um ou mais contextos.

Os instrumentos analisados no presente estudo, independentemente do tipo (diários, questionários, entrevistas), demonstraram maior reprodutibilidade “teste-reteste” do que validade. Em alguns instrumentos, os coeficientes de reprodutibilidade e validade foram elevados ($\geq 0,70$). Entretanto, a maioria apresentou baixos níveis de reprodutibilidade, mas, sobretudo, de validade ($\leq 0,50$), especialmente quando o “padrão-ouro” foi uma medida subjetiva. Poucas alterações foram verificadas nas evidências de validade ao considerar apenas os “padrões-ouro” mais robustos (acelerômetros, DLW, monitores de frequência cardíaca), sugerindo que a maioria dos instrumentos analisados mensura com limitações a atividade física dos adolescentes.

Em geral, maiores níveis de reprodutibilidade e de validade foram encontrados nos diários, se comparados aos questionários e entrevistas estruturadas. Contudo, há questionários que também alcançaram evidências aceitáveis de validade. Outros estudos são necessários para avaliar as evidências de reprodutibilidade e validade das entrevistas estruturadas.

Os instrumentos que foram testados em adolescentes brasileiros demonstraram níveis moderados a elevados de reprodutibilidade “teste-reteste” e baixos níveis de validade. Novos estudos precisam ser desenvolvidos para testar outros instrumentos, particularmente utilizando uma medida objetiva de atividade física como “padrão-ouro”.

Não há um instrumento que seja adequado a todos os estudos que envolvem medida de atividade física em adolescentes. Escolher o instrumento mais apropriado não depende somente dos níveis de reprodutibilidade e validade, mas também de outros fatores como as características da amostra, os objetivos do estudo, recursos disponíveis, os domínios e as dimensões da atividade física que serão mensurados e o período de referência da medida.

Resumo

Realizou-se revisão sistemática de estudos de reprodutibilidade e validade de instrumentos de medida da atividade física do tipo self-report – medidas subjetivas, em adolescentes (10-19 anos). Buscas foram realizadas em bases de dados (MEDLINE, PsycInfo, SportsDiscus, Scopus, Web of Science, SciELO, Lilacs) e nas referências dos artigos localizados. Sessenta e seis estudos atenderam aos critérios de inclusão. A maioria deles foi realizada em países da América do Norte, apenas cinco no Brasil. Identificaram-se 52 instrumentos diferentes: 42 questionários, 6 diários e 4 entrevistas. A reprodutibilidade “teste-reteste” variou de 0,20 a 0,98; a maioria (28/50) dos coeficientes apresentou valores < 0,70. Os coeficientes de validade apresentaram ampla variação (-0,13 a 0,88), sendo a maioria deles (64/84) ≤ 0,50. Apenas três instrumentos apresentaram correlações ≥ 0,70. Diversos instrumentos foram testados em adolescentes, especialmente questionários. Em geral, tais instrumentos demonstraram melhor reprodutibilidade “teste-reteste” do que validade.

Atividade Motora; Adolescente; Reprodutibilidade dos Testes; Validade dos Testes; Métodos

Colaboradores

J. C. Farias Júnior participou de todas as etapas do estudo, incluindo a busca e revisão dos estudos, análise e a redação do manuscrito. A. S. Lopes revisou e incluiu modificações à versão final do manuscrito. A. A. Florindo e P. C. Hallal participaram da estruturação dos métodos, análise dos resultados e da revisão crítica do manuscrito.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa outorgada para a execução do estudo e aos autores dos artigos revisados, pelas informações fornecidas, em especial àqueles que enviaram as cópias eletrônicas dos artigos.

Referências

- Dumith SC. Physical activity in Brazil: a systematic review. *Cad Saúde Pública* 2009; 25 Suppl 3: S415-26.
- Hallal PRC, Dumith SC, Bastos JP, Reichert FF, Siqueira FV, Azevedo MR. Evolução da pesquisa epidemiológica em atividade física no Brasil: revisão sistemática. *Rev Saúde Pública* 2007; 41:453-60.
- Baranowski T. Validity and reliability of self-report measures of physical activity: an information-processing perspective. *Res Q Exerc Sport* 1988; 59:314-27.
- Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. *J Appl Physiol* 2008; 105:977-87.
- Dollman J, Okely AD, Hardy L, Timperio A, Salmon J, Hills AP. A hitchhiker's guide to assessing young people's physical activity: deciding what method to use. *J Sci Med Sport* 2009; 12:518-25.
- Bouchard C, Shephard RJ. Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Shephard RJ, Stephens T, editors. *Physical activity, fitness and health: international proceedings and consensus statement*. Toronto: Human Kinetics Publishers; 1994. p. 11-23.
- Caspersen CJ, Poweell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985; 100:126-31.
- Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sports Med* 2001; 31:439-54.
- Kohl HW, Fulton JE, Caspersen CJ. Assessment of physical activity among children and adolescents: a review and synthesis. *Prev Med* 2000; 31:S54-76.
- Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. *Arch Dis Child* 2008; 93:614-9.
- Saris WHM. The assessment and evaluation of daily physical activity in children. A review. *Acta Paediatr Scand Suppl* 1985; 318:37-48.
- Armstrong N, Welsman JR. The physical activity patterns of European youth with reference to methods of assessment. *Sports Med* 2006; 36: 1067-86.
- Pate RR. Physical activity assessment in children and adolescents. *Crit Rev Food Sci Nutr* 1993; 33:321-6.
- Sallis JF. Self-report measures of children's physical activity. *J Sch Health* 1991; 61:215-9.
- Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport* 2000; 71(2 Suppl): S1-14.
- Crocker PRE, Holowachuk DR, Kowalski KC. Feasibility of using the tritrac motion sensor over a 7-day trial with older children. *Pediatr Exerc Sci* 2001; 13:70-81.
- Florindo AA, Romero A, Peres SV, Silva MV, Slater B. Desenvolvimento e validação de um questionário de avaliação da atividade física para adolescentes. *Rev Saúde Pública* 2006; 40:802-9.
- Gao S, Harnack L, Schmitz K, Fulton J, Lytle L, Coervering PV. Reliability and validity of a brief tool to measure children's physical activity. *J Phys Act Health* 2006; 3:415-22.
- Janz K, Witt J, Mahoney LT. The stability of children's physical activity as measured by accelerometry and self-report. *Med Sci Sports Exerc* 1995; 27:1326-32.
- Koo MM, Rohan TE. Comparison of four habitual physical activity questionnaire in girls aged 7-15 yr. *Med Sci Sports Exerc* 1999; 31:421-7.
- Kowalski KC, Crocker PRE, Faulkner RA. Validation of the physical activity questionnaire for older children. *Pediatr Exerc Sci* 1997; 9:174-86.
- McMurray RG, Ring KB, Treuth MS, Welk GJ, Pate RR, Schmitz KH, et al. Comparison of two approaches to structured physical activity surveys for adolescents. *Med Sci Sports Exerc* 2004; 36: 2135-43.
- Ridley K, Dollman J, Olds T. Development and validation of a computer delivered physical activity questionnaire (CDPAQ) for children. *Pediatr Exerc Sci* 2001; 13:35-46.
- Sallis JF, Buono MJ, Roby JJ, Micalo FG, Nelson JA. Seven-day recall and other physical activity self-reports in children and adolescents. *Med Sci Sports Exerc* 1993; 25:99-108.
- Welky GJ, Dziewaltowski DA, Hill JL. Comparison of the computerized ACTIVITYGRAM instrument and the previous day physical activity recall for assessing physical activity in children. *Res Q Exerc Sport* 2004; 75:370-80.
- Allor KM, Pivarnik JM. Stability and convergent validity of three physical activity assessments. *Med Sci Sports Exerc* 2001; 33:671-6.
- Garcia AW, George TR, Coviak C, Antonakos C, Pender NJ. Development of the child/adolescent activity log: a comprehensive and feasible measure of leisure-time physical activity. *Int J Behav Med* 1997; 4:323-38.
- Kemper HCG, Bakker I, Twisk JWR, van Mechelen W. Validation of a physical activity questionnaire to measure the effect of mechanical strain on bone mass. *Bone* 2002; 30:799-804.
- Nahas MVN, Barros MVG, Florindo AA, Farias Júnior JC, Hallal PC, Konrad L, et al. Reprodutibilidade e validade do questionário saúde na boa para avaliar atividade física e hábitos alimentares em escolares do ensino médio. *Rev Bras Ativ Fís Saúde* 2007; 12:12-20.
- Sallis JF, Strikmiller PK, Harsha DW, Feldman HA, Ehlinger S, Stone EJ. Validation of interviewer and self-administered physical activity checklists for fifth grade students. *Med Sci Sports Exerc* 1996; 28:840-51.
- Scerpella TA, Tuladhar P, Kanaley JA. Validation of the Godin-Shephard Questionnaire in prepubertal girls. *Med Sci Sports Exerc* 2002; 34:845-50.
- Simons-Morton BG, Taylor WC, Huang IW. Validity of the physical activity interview and caltrac with preadolescent children. *Res Q Exerc Sport* 1994; 65:84-8.

33. Verheul AM, Prins AN, Kemper HGC, Kardinaal AFM, Erp-Baart MJV. Validation of a weight-bearing physical activity questionnaire in a study of bone density in girls and women. *Pediatr Exerc Sci* 1998; 10:38-47.
34. Welky GJ, Wickel E, Peterson M, Heitzler CD, Fulton JE, Potter LD. Reliability and validity of questions on the youth media campaign longitudinal survey. *Med Sci Sports Exerc* 2007; 39:612-22.
35. Weston AT, Petosa R, Pate RR. Validation of an instrument for measurement of physical activity in youth. *Med Sci Sports Exerc* 1997; 29:138-43.
36. Brener ND, Collins JL, Kann L, Warren CW, Williams BI. Reliability of the youth behavior survey questionnaire. *Am J Epidemiol* 1995; 141:575-80.
37. Brener ND, Kann L, McManus T, Kinchen SA, Sundberg EC, Ross JG. Reliability of the 1999 Youth Risk Behavior Survey Questionnaire. *J Adolesc Health* 2002; 31:336-42.
38. Brown TD, Holland BV. Test-retest reliability of the self-assessed physical activity checklist. *Percept Mot Skills* 2004; 99:1099-102.
39. Crocker PRE, Bailey DA, Faulkner RA, Kowalski KC, McGrath R. Measuring general levels of physical activity: preliminary evidence for the physical activity questionnaire for older children. *Med Sci Sports Exerc* 1997; 29:1344-9.
40. Farias Júnior JC, Pires MC, Lopes AS. Reprodutibilidade de um questionário para o levantamento de informações sobre comportamentos relacionados à saúde em adolescentes. *Rev Bras Ciênc Mov* 2002; 10:43-8.
41. Gilmer MJ, Speck BJ, Bradley C, Harrel JS, Belyea M. The youth health survey: reliability and validity on an instrument for assessing cardiovascular health habits in adolescents. *J Sch Health* 1996; 66:106-11.
42. Kriska AM, Knowler WC, LaPorte RE, Drash AL, Wing KR, Blair SN, et al. Development of questionnaire to examine relationship of physical activity and diabetes in Pima Indians. *Diabetes Care* 1990; 13:401-11.
43. Zullig KJ, Pun S, Patton JM, Ubbes VA. Reliability of the 2005 middle school youth risk behavior survey. *J Adolesc Health* 2006; 39:856-69.
44. Anderson CB, Hagströmer M, Yngve A. Validation of the PDPAR as an adolescent diary: effects of accelerometer cut points. *Med Sci Sports Exerc* 2005; 37:1224-30.
45. Arvidsson D, Slinde F, Hulthén L. Physical activity questionnaire for adolescents validated against doubly labelled water. *Eur J Clin Nutr* 2005; 59: 376-83.
46. Bratteby L-E, Sandhagen B, Fan H, Samuelson G. A 7-day activity diary for assessment of daily energy expenditure validated by the doubly labelled water method in adolescents. *Eur J Clin Nutr* 1997; 51:585-91.
47. Cradock AL, Wiecha JL, Peterson KE, Sobol AM, Colditz GA, Gortmaker SL. Youth recall and tri-trac accelerometer estimates of physical activity levels. *Med Sci Sports Exerc* 2004; 36:525-32.
48. Ekelund U, Neovius M, Linné Y, Rössner S. The criterion validity of a last 7-day physical activity questionnaire (SAPAQ) for use in adolescents with a wide variation in body fat: the Stockholm Weight Development Study. *Int J Obes (Lond)* 2006; 30:1019-21.
49. Ekelund U, Yngvel A, Sjostrom M. Total daily energy expenditure physical activity in adolescents by two different methods. *Scand J Med Sci Sports* 1999; 9:257-64.
50. Kowalski KC, Crocker PRE, Kowalski NP. Convergent validity of the physical activity questionnaires for adolescents. *Pediatr Exerc Sci* 1997; 9:342-52.
51. Larsson C, Westerterp KR, Johansson GJ. Validity of reported energy expenditure and energy and protein intakes in Swedish adolescent vegans and omnivores. *Am J Clin Nutr* 2002; 75:268-74.
52. Moore JB, Hanes Jr. JC, Barbeau P, Gutin B, Treviño RP, Yin Z. Validation of the Physical Activity Questionnaire for Older Children in children of different races. *Pediatr Exerc Sci* 2007; 19:6-19.
53. Murphy JK, Alpert BS, Dupaul LM, Willey ES, Walker SS, Nanney GC. The validity of children's self-report of physical activity: a preliminary study. *J Hum Hypertens* 1990; 4:130-2.
54. Pate RR, Ross R, Dowda M, Trost SG, Sirard J. Validation of a 3-day physical activity recall instrument in female youth. *Pediatr Exerc Sci* 2003; 15:257-65.
55. Ridder CM, Kemper HCG, Bertensy MJM, van Gamen ACL, Ras E, Voogd J. Concurrent validity of a weight-bearing activity questionnaire in prepubertal and pubertal girls and boys. *Ann Hum Biol* 2002; 29:237-46.
56. Rodriguez G, Béghin L, Michaud L, Moreno LA, Turck D, Gottrand F. Comparison of the TriTrac-R3D accelerometer and self-report activity diary with heart-rate monitoring for the assessment of energy expenditure in children. *Br J Nutr* 2002; 87:623-31.
57. Slinde F, Arvidsson D, Sjoeborg A, Rossander-Hulthen L. Minnesota leisure time activity questionnaire and doubly labeled water in adolescents. *Med Sci Sports Exerc* 2003; 35:1923-8.
58. Trost SG, Marshall AL, Miller R, Hurley JT, Hunt JA. Validation of a 24-h physical activity recall in indigenous and non-indigenous Australian adolescents. *J Sci Med Sport* 2007; 10:428-35.
59. Trost SG, Ward DS, McGraw B, Pate RR. Validity of the previous day physical activity recall (PDPAR) in fifth-grade children. *Pediatr Exerc Sci* 1999; 11:341-5.
60. Wallace JP, McKenzie TL, Nader PR. Observed vs. recalled exercise behavior: a validation of a seven day exercise recall for boys 11 to 13 years old. *Res Q Exerc Sport* 1985; 56:161-5.
61. Argiropoulou EC, Michaloupoulou M, Aggeloussis N, Avgerinos A. Validity and reliability of physical activity measures in Greek high school age children. *J Sport Sci Med* 2004; 3:147-59.
62. Bouchard C, Tremblay A, Leblanc C, Lortie G, Sauard R, Theriault G. A method to assess energy expenditure in children and adults. *Am J Clin Nutr* 1983; 37:461-7.

63. Eisenmann JC, Milburn N, Jacobsen L, Moore SJ. Reliability and convergent validity of the Godin leisure-time exercise questionnaire in rural 5th-grade school-children. *Journal of Human Movement Studies* 2002; 43:135-49.
64. Lee K, Trost SG. Validity and reliability of the 3-day physical activity recall. *Res Q Exerc Sport* 2005; 76:101-6.
65. Mota J, Santos P, Guerra S, Ribeiro JC, Duarte JA, Sallis JF. Validation of a physical activity self-report questionnaire in a portuguese pediatric population. *Pediatr Exerc Sci* 2002; 14:259-76.
66. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. *Arch Pediatr Adolesc Med* 2001; 155:554-9.
67. Troped PJ, Wiecha JL, Fragala MS, Matthews CE, Finkelstein DM, Kim J, et al. Reliability and validity of YRBS physical activity items among middle school students. *Med Sci Sports Exerc* 2007; 39: 416-25.
68. Ching PLYH, Dietz WH. Reliability and validity of activity measures in preadolescents girls. *Pediatr Exerc Sci* 1995; 7:389-99.
69. Guedes DP, Lopes CC, Guedes JERP, Stanganelli LC. Reprodutibilidade e validade do questionário Baecke para avaliação da atividade física habitual em adolescentes. *Rev Port Ciênc Desporto* 2006; 6:265-74.
70. McMurray RG, Harrell JS, Bradley CB, Webb JP, Goodman EM. Comparison of a computerized physical activity recall with a triaxial motion sensor in middle-school youth. *Med Sci Sports Exerc* 1998; 30:1238-45.
71. Telford A, Salmon J, Jolley D, Crawford D. Reliability and validity of physical activity questionnaires for children: the children's leisure activities study survey (CLASS). *Pediatr Exerc Sci* 2004;16:64-78.
72. Ridley K, Olds TS, Hill A. The Multimedia Activity Recall for Children and Adolescents (MARCA): development and evaluation. *Int J Behav Nutr Phys Act* 2006; 3:10.
73. Treuth MS, Hou N, Young DR, Maynard M. Validity and reliability of the fels physical activity questionnaire for children. *Med Sci Sports Exerc* 2005; 37:488-95.
74. Hernández B, Gortmaker SL, Laird NM, Colditz GA, Parra-Cabrera S, Peterson KE. Validez y reproducibilidad de um questionario de actividad e inactividad física para escolares de la ciudad de México. *Salud Pública Méx* 2000; 42:315-23.
75. Biddle S, Sallis JF, Cavill NA. Young and active? Young people and health enhancing physical activity: evidence and implication. London: Health Education Authority; 1998.
76. Aaron DJ, Kriska AM, Dearwater SR, Cauley JA, Metz KF, LaPorte RE. Reproducibility and validity of an epidemiologic questionnaire to assess part year physical activity in adolescents. *Am J Epidemiol* 1995; 142:191-201.
77. Guedes DP, Lopes CC, Guedes JERP. Reprodutibilidade e validade do questionário internacional de atividade física em adolescentes. *Rev Bras Med Esporte* 2005; 11:151-8.
78. Booth ML, Okely AD, Chey T, Bauman A. The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HSBC) survey: a population study. *Br J Sports Med* 2001; 35:263-7.
79. Wong SL, Leatherdale ST, Manske SR. Reliability and validity of a school-based physical activity questionnaire. *Med Sci Sports Exerc* 2006; 38: 1593-600.
80. Philippaerts RM, Matton L, Wijndaele K, Balduck AL, De Bourdeaudhuij I, Lefevre J. Validity of a physical activity computer questionnaire in 12- to 18-year-old boys and girls. *Int J Sports Med* 2006; 27:131-6.
81. Barbosa N, Sanchez C, Vera J, Perez W, Thalabard JC, Rieu M. A physical activity questionnaire: reproducibility and validity. *Journal of Sports Science and Medicine* 2007; 6:505-18.
82. Booth ML, Okely AD, Chey T, Bauman A. The reliability and validity of the adolescent physical activity recall questionnaire. *Med Sci Sports Exerc* 2002; 34:1986-95.
83. Patterson P. Reliability, validity, and methodological response to the assessment of physical activity via self-report. *Res Q Exerc Sport* 2000; 71 (2 Suppl):S15-20.
84. Shephard R. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 2004; 37:197-206.
85. Baumgartner TA, Chung H. Confidence limits for intraclass reliability coefficients. *Measurement in Physical Education and Exercise Science* 2001; 5:179-88.
86. Reilly J, Penpraze V, Hislop J, Davies G, Grant S, Patton J. Objective measurement of physical activity and sedentary behaviour: review with new data. *Arch Dis Child*. 2008; 93:614-9.
87. Bland JM, Altman G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327:307-10.
88. Monteiro CA, Florindo AA, Claro RM, Moura EC. Validade de indicadores de atividade física e sedentarismo obtidos por inquérito telefônico. *Rev Saúde Pública* 2008; 42:575-81.

Recebido em 14/Jan/2010

Versão final reapresentada em 30/Mai/2010

Aprovado em 11/Jun/2010