

# Validity and Reliability of Field-Based Measures for Assessing Movement Skill Competency in Lifelong Physical Activities: A Systematic Review

Ryan M. Hulteen<sup>1</sup> · Natalie J. Lander<sup>2</sup> · Philip J. Morgan<sup>1</sup> · Lisa M. Barnett<sup>2</sup> · Samuel J. Robertson<sup>3</sup> · David R. Lubans<sup>1</sup>

Published online: 15 July 2015  
© Springer International Publishing Switzerland 2015

## Abstract

**Background** It has been suggested that young people should develop competence in a variety of ‘lifelong physical activities’ to ensure that they can be active across the lifespan.

**Objective** The primary aim of this systematic review is to report the methodological properties, validity, reliability, and test duration of field-based measures that assess movement skill competency in lifelong physical activities. A secondary aim was to clearly define those characteristics unique to lifelong physical activities.

**Data Sources** A search of four electronic databases (Scopus, SPORTDiscus, ProQuest, and PubMed) was conducted between June 2014 and April 2015 with no date restrictions.

**Study Selection** Studies addressing the validity and/or reliability of lifelong physical activity tests were reviewed. Included articles were required to assess lifelong physical

activities using process-oriented measures, as well as report either one type of validity or reliability.

**Study Appraisal and Synthesis Methods** Assessment criteria for methodological quality were adapted from a checklist used in a previous review of sport skill outcome assessments.

**Results** Movement skill assessments for eight different lifelong physical activities (badminton, cycling, dance, golf, racquetball, resistance training, swimming, and tennis) in 17 studies were identified for inclusion. Methodological quality, validity, reliability, and test duration (time to assess a single participant), for each article were assessed. Moderate to excellent reliability results were found in 16 of 17 studies, with 71 % reporting inter-rater reliability and 41 % reporting intra-rater reliability. Only four studies in this review reported test–retest reliability. Ten studies reported validity results; content validity was cited in 41 % of these studies. Construct validity was reported in 24 % of studies, while criterion validity was only reported in 12 % of studies.

**Limitations** Numerous assessments for lifelong physical activities may exist, yet only assessments for eight lifelong physical activities were included in this review. Generalizability of results may be more applicable if more heterogeneous samples are used in future research.

**Conclusion** Moderate to excellent levels of inter- and intra-rater reliability were reported in the majority of studies. However, future work should look to establish test–retest reliability. Validity was less commonly reported than reliability, and further types of validity other than content validity need to be established in future research. Specifically, predictive validity of ‘lifelong physical activity’ movement skill competency is needed to support the assertion that such activities provide the foundation for a lifetime of activity.

---

✉ David R. Lubans  
David.Lubans@newcastle.edu.au

<sup>1</sup> Priority Research Centre in Physical Activity and Nutrition, University of Newcastle, Callaghan, NSW, Australia

<sup>2</sup> School of Health and Social Development, Deakin University, Burwood, VIC, Australia

<sup>3</sup> Institute for Sport, Exercise and Active Living, Victoria University, Footscray, VIC, Australia

## Key Points

Lifelong physical activities are typically performed individually or in small groups, involve minimal structure and minimal physical contact, are characterized by varying levels of intensity and competitiveness, and may be easily carried into adulthood and old age.

Additional research is needed to establish the validity and reliability of lifelong physical activity movement skill tests for activities not included in this review, such as yoga, Pilates, tai chi, aerobics, and running.

Future research would benefit from determining the predictive validity of competency in lifelong physical activities to ascertain the strength and direction of association between competency levels and future physical activity.

## 1 Introduction

Developing adequate movement skill competency across a broad range of activities is important for individuals of all ages [1–3], and competency in a range of fundamental movement skills (FMS) in childhood has been found to be a predictor of physical activity in adolescence [4]. Movement skills are often learned and developed throughout childhood [5–7], initially in the form of FMS, of which there are three types: locomotor (i.e., running, jumping), object control (i.e., catching, kicking), and stability (i.e., balancing, twisting) [8]. If children fail to develop competency in FMS [9, 10], they may find it difficult to learn and master more refined movement skills, such as sport-specific skills (i.e., pitching a ball, serving in tennis) [7].

Previous movement skill competence theory [7, 11] posits that individuals ascend a hypothetical mountain of motor development, whereby more advanced movement acquisition is dependent upon the foundation established in the previous level. The proposed levels of movement skill acquisition are (1) reflexive, (2) preadapted, (3) fundamental motor patterns, (4) specialized sports skills [11], and (5) skillful [7]. These models are based on the premise that individuals cannot be physically active throughout the lifespan without achieving proficiency in FMS. However, some lifelong physical activities do not require a foundation in FMS that are often assessed, meaning children who are not competent in FMS may alternatively perform lifelong physical activities to be physically active. As such, it has been suggested that young people need to be exposed

to, and develop competency in, a range of movement skills associated with ‘lifelong physical activities’ that can be easily carried into adulthood [12–17].

Schools may present a possible setting for learning and testing competency in lifelong physical activities, as they may have access to personnel and resources, such as qualified teachers, equipment, space, and the ability through physical education to provide exposure to these activities [14, 16, 18, 19]. A noted decline in physical activity occurs during adolescence [20]; thus, this may also be a critical period in which individuals should learn and develop competency in a range of lifelong physical activities. Indeed, lifelong activities learned at this time may have health benefits both at the time they are learned and later on in adult years [14].

Although a variety of definitions and alternative terminology for lifelong physical activities (i.e., lifetime [14], lifestyle physical activities [13, 16, 21]) have emerged in the literature [13, 22–24], different characteristics appear regarding what defines a lifelong physical activity, which consequently makes identifying and promoting these activities difficult. Of note, the term ‘lifelong physical activity’ can also be used to describe how an individual can be physically active across the lifespan. It is proposed that the term ‘lifelong physical activity’ will only be used to describe a subset of physical activities defined as those sports and leisure-time activities typically performed individually or in small groups (typically four or fewer people) that involve minimal structure, avoid physical contact, are characterized by varying levels of intensity and competition, and, importantly, may be easily carried into adulthood and old age. Examples of lifelong physical activities that fit this definition include aerobics, badminton, cycling, dance, golf, Pilates, racquetball, resistance training, running, swimming, tai chi, tennis, and yoga.

Many team sports, such as basketball, hockey, and soccer can be played throughout adulthood, but do not fit the definition of a lifelong activity due to the number of participants and the higher levels of organization required [14]. In addition, due to the physical contact involved, many team sports have higher incidences of injury (number of injuries/1000 occurrences), such as soccer (64.4/1000), rugby (95.7/1000), and hockey (62.6/1000) [25–27]. In comparison, popular lifelong physical activities, such as tennis (23.1/1000), resistance training (11.9/1000), and swimming (6.1/1000) [25–27] have considerably lower injury rates.

One characteristic of lifelong physical activities included in previous definitions, but not in this study, is the use of minimal equipment. While this may be true for most lifelong activities, there are some notable exceptions, including golf and resistance training. Although golf could be played with only two clubs (i.e., an iron and a putter),

this is neither ideal nor representative of how golf is usually played. Similarly, resistance training is often performed with free weights (e.g., barbells and dumbbells) and equipment (e.g., cables and pulleys) typically found in a health club, yet it can also be performed using only body weight exercises (e.g., squats, lunges, push-ups). Regardless of equipment, these two activities are undoubtedly lifelong activities, as shown by their high levels of participation amongst individuals of all ages [28]. Although equipment considerations may be important in performing an activity, the inclusion of this characteristic in the definition of lifelong physical activities would exclude relevant lifelong physical activities from this review.

The assessment of movement skills (e.g., FMS, sport-specific, lifelong) is vital for informing individuals of their competency levels, as well as informing teachers and researchers of potential movement skill deficiencies in a population, so programs or interventions can be designed and implemented [29]. Movement skills are commonly assessed through product or process measures [8]. Product measures quantify outcomes [8], which are expressed, for example, in terms of how fast a ball is thrown (i.e., speed), time it takes to swim 100 m, or distance a soccer ball is kicked. Product measures are quick and easy to assess and interpret, but they cannot be used to determine how an outcome was achieved [29]. Conversely, process-oriented measurement is concerned with the qualitative characteristics that describe successful movement patterns [8] and allow movement component deficiencies to be more easily identified and corrected. The ability to correct individuals on specific components of a movement may help prevent injury [30], may help contribute to an individual's feeling of competence, and can enable the identification of specific skill components to be addressed in future interventions to enhance performance proficiency. For example, when an individual is introduced to a new activity, such as weightlifting, the technique, as opposed to the outcome (i.e., amount of weight lifted), is more important for the safety of the individual [1]. Over time, if safe technique is practiced and utilized, strength gains may be achieved with reduced possibility of injury [1].

Previous reviews have examined the validity and reliability (alternatively called objectivity) of assessments in both FMS [31] and sport-specific skills [32]. No such review exists for the assessment of field-based measures (i.e., not taking place in a laboratory setting, but rather in, for example, a school or community sporting ground) of lifelong physical activities. Given the widespread use [33–35] and previous success of process-oriented skill assessment in FMS (e.g., Test of Gross Motor Development-2 [TGMD-2]) [36], an analysis of measurement properties of current assessments examining qualitative aspects of lifelong physical activities is warranted. Therefore, the

purpose of this systematic review is to review the methodological properties, validity, reliability, and test duration of current field-based measures to assess movement skill competency in lifelong physical activities, as well as clearly define the characteristics unique to lifelong physical activities.

## 2 Methods

### 2.1 Search Strategy

A systematic review of four search engines (PubMed, Scopus, ProQuest, and SPORTDiscus) was conducted, focused on field-based measures of lifelong activities. No time restrictions were applied when searching for articles. Searches conducted in the individual databases included various combinations of the following terms: 'reliability' OR 'validity' AND 'fitness' OR 'physical activity' OR 'sport' OR 'motor' OR 'movement' OR 'skill' OR 'battery' OR 'instrument' OR 'qualitative' OR 'technique' OR 'components' OR 'criteria' OR 'measurement' OR 'test' OR 'assessment'. A secondary search for specific lifelong physical activities (aerobics, badminton, cycling, dance, golf, Pilates, racquetball, resistance training, running, swimming, tai chi, tennis and yoga) was performed. Additional articles were found by examining the reference lists of included articles. After the initial searches, the titles and abstracts of all relevant articles were assessed. If the articles were deemed appropriate, then a full-text review was performed, and the application of inclusion and exclusion criteria allowed for further evaluation of included review articles.

### 2.2 Inclusion Criteria

Two authors independently assessed articles for inclusion in the study. If an agreement could not be reached, a third author reviewed and made the final decision on whether the article should be included. The criteria for inclusion in the study were as follows: (1) articles must have been peer reviewed; (2) full abstract, article, and reference list must be present; (3) articles must report at least one lifelong physical activity movement skill; and (4) article must report at least one aspect of validity or reliability relating to the movement skill. If a movement battery was used to test multiple skills, then the skill was only included if the skill and corresponding validity or reliability information could be extracted.

As assessments examining skill proficiency should display adequate measurement properties, it is important to consider the validity (e.g., content, construct, and/or criterion) of the measure. Content validity is concerned with

whether a test is a measure of all skills relevant to a particular activity [37, 38]. For example, it could be assumed that the content validity of a tennis assessment is higher in a test that assesses the forehand, backhand, volley, and serve, as opposed to a test that examines just the forehand. Construct validity is a measure of whether a test can measure a quality or attribute that cannot be operationalized. It consists of discriminative (ability to assess performers of different ability by another measure) and convergent (relation of a test with another measure of the same construct or associated measures) validity [38, 39]. Finally, criterion validity refers to the ability of a test to show agreement with a ‘gold standard’ or external measure. Criterion validity can also constitute concurrent (relating score with a ranking in an alternative measure) or predictive (relationship of a score to a future performance) validity [38].

Three main types of reliability were reported for the studies included in this review. Inter-rater reliability was defined as the agreement between two or more raters on an assessment/score [39]. Intra-rater reliability was defined as the level of agreement of a single observer on multiple assessments/scores [39]. Finally, test–retest reliability is

defined as the level of consistency over two or more rounds of testing [39].

### 2.3 Exclusion Criteria

Studies were excluded if they met any of the following criteria: (1) the activity did not fit the definition of a life-long activity; (2) insufficient information on validity and/or reliability was reported; (3) the skill was assessed via use of a product measure; (4) the qualitative criteria for measuring the skill were not clearly defined; or (5) articles were not reported in English.

### 2.4 Assessment of Study Quality

Two authors independently reviewed all included articles for study quality (see Table 1) based upon five criteria adapted from a risk-of-bias assessment in a previously published review on sport-related skill outcomes [32]. The five criteria by which articles were assessed included (1) sample size, reported as the number of participants used specifically for establishing the validity and/or reliability of the skill test; (2) participant details, which included age,

**Table 1** Risk of bias

Reference	Activity	Sample size	Participant details	Practice session	Stability of conditions	Time between assessments
Toriola et al. [49]	Badminton	116	+/-	-	-	NA
Wang et al. [54]	Badminton	30	+	-	-	NA
Ducheyne et al. [46]	Cycling	93	+/-	-	+/-	NA
		20 (reliability)				
Macarthur et al. [52]	Cycling	20	+/-	-	+/-	3 months
Angioi et al. [50]	Dance	6	+	+	-	NA
Chatfield [47]	Dance	41	+	-	+	Immediately
Krasnow and Chatfield [44]	Dance	20	+/-	+	+/-	8 weeks
Barnett et al. [48]	Golf	43	+	+	+	5 days
		19–28 (reliability)				
Lam and Zhang [43]	Racquetball	131	+	+ <sup>a</sup>	+	Within 1 week of first test
		20 (reliability)				
Barnett et al. [59]	Resistance training	12	+/-	-	+	NA
Lubans et al. [45]	Resistance training	63	+/-	-	+	7 days
Myer et al. [42]	Resistance training	-	-	-	-	-
Erbaugh [56]	Swimming	57	+/-	+	+	1 week
		22 (test–retest)				
Zetou et al. [55]	Swimming	46	+/-	-	+	6 weeks
Buszard et al. [51]	Tennis	28	+	-	+	-
Farrow et al. [57]	Tennis	23	+/-	+	+	6 weeks
Messick [53]	Tennis	60	+	-	-	-

NA not applicable, + indicates fully reported, - indicates not reported, +/- indicates partially reported

<sup>a</sup> Familiarization session was optional

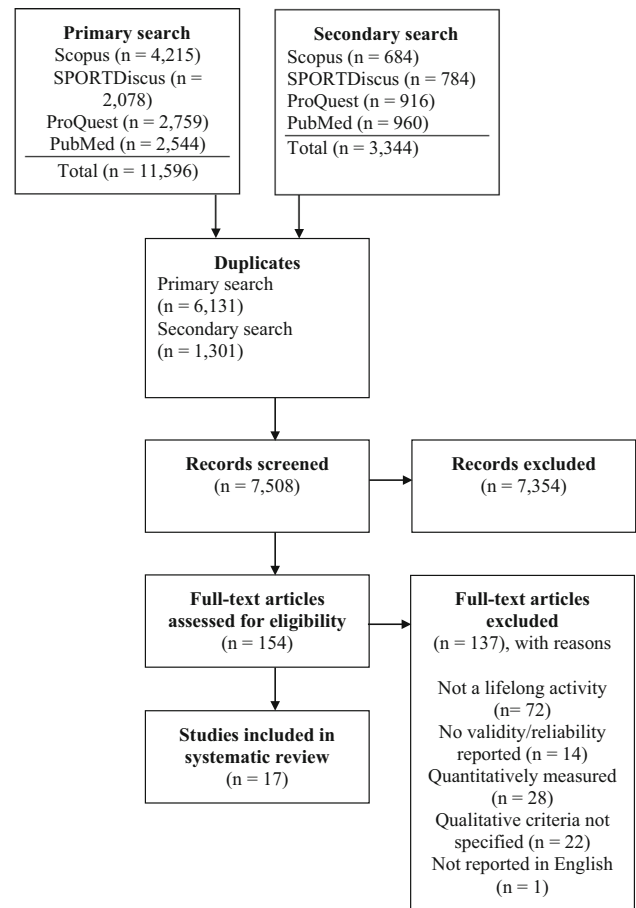
sex, number of participants, and ability level; (3) whether participants were allowed to practice the tested skills before the official assessment (practice session information was simply reported as having occurred or not); (4) testing environment, including the equipment remaining the same throughout the entire testing process, which was reported as yes or no, or a partial report was given if the stability of conditions can be implied due to study design; (5) reported amount of time between assessments, if applicable. Along with study quality, authors extracted validity and reliability results from each article. As general group associations are determined using correlation coefficients ( $r$ ) and intraclass correlation coefficient (ICC), values were classified as follows:  $<0.4$  was rated as poor,  $\geq 0.4$  to  $<0.8$  was moderate, and  $\geq 0.8$  as excellent [39, 40]. As the  $\kappa$  coefficient is a measure of exact agreement between raters, a slightly modified scale was used:  $>0.01$  and  $\leq 0.2$  was rated as poor,  $>0.2$  and  $\leq 0.4$  was rated as fair,  $>0.4$  and  $\leq 0.6$  was moderate,  $>0.6$  and  $\leq 0.8$  was good, and  $>0.8$  and  $\leq 1.0$  was excellent [41]. If authors could not agree at any point during the data extraction phase, a third author made the final decisions on study quality and validity/reliability extraction.

### 3 Results

Preliminary search results identified 7508 articles; however, after examining titles and abstracts, 154 full-text articles were retrieved and reviewed for eligibility for inclusion in this review. Reasons for exclusion of search results can be viewed in Fig. 1. After inclusion/exclusion criteria were applied to the full-text articles, 17 met all criteria for inclusion into this review. These 17 articles consisted of eight different lifelong physical activities, including resistance training (three), badminton (two), tennis (three), cycling (two), racquetball (one), swimming (two), golf (one), and dance (three) articles. More specific information related to the skills tested, equipment needed, and the sample used in each study can be viewed in Table 2.

#### 3.1 Risk of Bias

Overall, relative to the study type and design, the sample sizes ranged from small ( $n = 6$ ) to very large ( $n = 131$ ). One study only established content validity and did not report a sample size [42]. Only 12 % of studies had a sample size greater than 100, while 47 % of studies involved small sample sizes of 30 participants or fewer. When reporting participants' details (i.e., sex, age, level of experience, and number of participants), only seven studies adequately reported these details, while the remaining ten



**Fig. 1** Results of systematic literature search

were missing at least one criterion. As previously reported by Robertson et al. [32], the ability level of study participants/cohorts is commonly not reported in studies, and this holds true for the current review as this detail went unrecorded more than any other participant detail ( $n = 8$ ). Six of the 17 included articles allowed participants to practice the studied skills before the official test was undertaken. However, it should be noted that one study [43] had an optional practice session; thus, we were not able to determine whether all participants had practiced. Nine studies reported keeping testing conditions the same between assessments (i.e., environment and equipment), while the remaining eight studies either made no mention of keeping testing conditions the same or the stability of conditions could not be deduced due to study design.

#### 3.2 Validity

Content validity was the most commonly reported validity for studies in this review. A total of 41 % of studies cited content validity, and all of these studies used some type of expert panel to establish the relevant skills/domains to be



**Table 2** Skills tested, equipment used, and participants involved in skill tests

Reference	Activity	Skill(s) tested	Equipment	Sample
Toriola et al. [49]	Badminton	Badminton overhead serve	Badminton court, racquet, shuttlecock, net, rope	Mixed sex physical education major college students
Wang et al. [54]	Badminton	Overhead badminton forehand strike	Badminton racquet, shuttlecock, video camera	Beginner level players aged 16–19 years
Ducheyne et al. [46]	Cycling	Walk with bicycle, mount bicycle and ride, look left and right while riding, ride in straight line, cycle in a circle, cycle one-handed in a circle, slalom, cycling on sloping surface, signal left and right while riding, brake to a controlled stop, dismount bicycle	Asphalt surface, bicycle, cones, boards, traffic signs, ramp, floor markers	Grade 4 primary school children
Macarthur et al. [52]	Cycling	Riding in a straight line, coming to a complete stop, shoulder checking before a left turn	Chalk, stop sign, pylon	Year 4 primary school children
Angioi et al. [50]	Dance	Control of movements, spatial skills, accuracy of movements, technique, dynamics, timing, and rhythmical accuracy, performance qualities, overall performance	Video camera	Professional contemporary dancers
Chatfield [47]	Dance	Skill, space, time, energy, phrasing, and presence	Soundtrack, video camera	Non-dancers, beginners, intermediates, advanced, and professional dancers
Krasnow and Chatfield [44]	Dance	Full-body involvement, body integration and connectedness, articulation of body segments, and movement skills	Video camera	Low-intermediate to advanced dance class college students
Barnett et al. [48]	Golf	Golf swing, golf putt	Foam golf ball, golf putter, golf club, cone, video camera	Children aged 6–10 years
Lam and Zhang [43]	Racquetball	Racquetball forehand, racquetball backhand, ceiling shot, wall rally	Standard racquetball court, racquets and racquetballs	Beginner level college students
Barnett et al. [59]	Resistance training	Body weight squat, push-up, lunge, suspended row, standing overhead press, front support with chest touch	Suspension bar/straps, anchor point (e.g. door), video camera	Year 7–10 school children
Lubans et al. [45]	Resistance training	Body weight squat, push-up, lunge, suspended row, standing overhead press, front support with chest touch	Suspension bar/straps, anchor point (e.g. door), video camera	Year 7–10 school children
Myer et al. [42]	Resistance training	Back squat	Cylindrical dowel	NA
Erbaugh [56]	Swimming	Water entry, front locomotion, back locomotion, breathing, kicking, diving, ring pick-up and retrieval	Pool, 6 × 6 platform, 12 × 3/4 bench, water rings, hoops, lane lines	Preschool children
Zetou et al. [55]	Swimming	Backstroke	Pool, pool buoy, video camera	Children aged 10–12 years
Buszard et al. [51]	Tennis	Tennis forehand	Small, medium and large tennis racquet, standard yellow, low compression green and very low compression red tennis ball, two video cameras	Beginner level children aged 6–8 years
Farrow et al. [57]	Tennis	Tennis forehand, tennis backhand	Tennis racquet, standard and modified 'red' tennis ball, regular and scaled tennis court with nets, two video cameras	Beginner level primary school children
Messick [53]	Tennis	Overhead tennis serve	Standard yellow tennis ball, tennis racquet, video camera	Mixed experience players, aged 9–19

NA not applicable

assessed. Two of the studies [43, 44] additionally used a literature review to further justify the inclusion of specific skills to allow for adequate content validity of their test.

Construct validity was reported in 24 % of studies. Of the studies reporting construct validity, three different statistical analyses were used. Lubans et al. [45] used a regression model, involving the total score of the resistance skill training battery and sex, and found that 39 % of variance could be explained by a muscular fitness score. Ducheyne et al. [46] established construct validity through factor analysis, which resulted in three factors being extracted from the skills test, including during-cycling skills, walking with the bicycle, and dismounting the bicycle. Discriminative validity was established for a test of dance and golf proficiency [47, 48]. An analysis of variance was used to test for group differences between ability level (i.e., non-dancers, beginners, intermediates, advanced, and professionals) and overall dance test scores. Alternatively, the golf assessment tested for differences in golf skill competency according to age (e.g., 6, 7/8, 9/10 year olds).

Only two studies tested for criterion-related validity. Toriola et al. [49] classified participants as low-skill (displaying less than 50 % of badminton service components) or high-skill (displaying more than 50 % of badminton service components) badminton players. Participants were then scored on a service test (i.e., quantitative test based on where the shuttlecock landed on the serve) while simultaneously being assessed by the judges on the quality of their movement. The results from these two assessments were then correlated, which yielded a low positive association for both low-skill ( $r = 0.04$ ) and high-skill ( $r = 0.06$ ) performers. These results indicate that the judges' process-oriented scoring of participants (i.e., quality of movement) could not sufficiently determine participants' scores on the overhead serve test (i.e., quantitative score). Similarly, process-oriented ratings on a racquetball skill battery [43] were used to assess the quality of participants' movements for eight different racquetball skills. This rating was then correlated to individuals' final standing in a racquetball tournament. This study revealed a higher relationship ( $r = -0.48$ ) compared with the badminton service test. A rank of one indicates the best player (i.e., high racquetball ability), whereas a score of ten would indicate the tenth best player in the tournament (i.e., less racquetball ability). Therefore, while criterion validity may provide important information in terms of predicting future performance or how a skill test compares to 'gold standards', the results of studies included in this review may show that more research should focus on improving and/or establishing criterion validity for use in process-oriented tests.

The validity results of included articles are displayed in Table 3. Six tests in this review failed to report any type of validity [50–55].

### 3.3 Reliability

All but one study [42] included in this review reported at least one type of reliability. Most common was inter-rater reliability ( $n = 12$ ). This was reported either as the percentage of agreement [53, 54, 56],  $r$  coefficient [44, 47, 49, 55, 56], ICC [46, 49–51], or a  $\kappa$  coefficient [52]. Intra-rater reliability was reported in 41 % of studies in a similar fashion as inter-rater reliability, with three studies reporting  $r$  coefficients [44, 47, 55], three reporting ICCs [43, 48, 50], and one study using percentage of agreement [53]. While most studies showed a high level of inter- and intra-rater agreement, one study [53] had questionable levels of inter-rater agreement (i.e., percentage of agreement below 80 %) for two of the six components assessing the overhead tennis serve.

Test–retest reliability was only reported in four studies [45, 48, 56, 57]. Of those studies reporting test–retest reliability, two studies reported this as an  $r$  coefficient [56, 57] and one as an ICC [48]. The fourth study reporting test–retest reliability was unique in that this was demonstrated through rank order repeatability (i.e., ability of participants to remain the same across multiple trials) and change in mean (i.e., change in score between trials of an individual as opposed to group differences and typical error [45, 58]). These statistics were unique to the resistance training battery identified in this review, and the authors of the paper were comparing differences between individuals, unlike other tests that compare group differences. Additionally, coefficient of variation [59] was used in another article assessing the resistance training battery to further show the reliability of the instrument. Two studies reported three different types of reliability statistics [45, 56], while all other studies reported either one or two reliability statistics. Overall, however, levels of reliability were moderate to excellent, with no ICC below 0.60,  $r$  below 0.67, and percent agreement below 69 %.

### 3.4 Test Duration

To the authors' knowledge, no published guidelines for determining adequate test duration exist. However, test duration has been used as one component of feasibility in a previous sport skill review [32]. Thus, duration to assess a single participant (independent of set-up time) was extracted for this review. Eight of the 17 articles reported time to assess a single participant in a skill test/battery [43,

**Table 3** Measurement properties

Reference	Sport	Validity	Reliability	Test duration
Toriola et al. [49]	Badminton	Criterion (concurrent $r = 0.04\text{--}0.06$ )	Inter-rater (Kendall coefficient of concordance = 0.68, 0.70, 0.78)	NR
Wang et al. [54]	Badminton	NR	Inter-rater (87.3–94.8 %)	First 5 min of session were recorded
Ducheyne et al. [46]	Cycling	Construct (factor analysis)	Inter-rater (ICC 0.75–0.98)	NR
Macarthur et al. [52]	Cycling	NR	Inter-rater ( $\kappa = 0.9\text{--}1.0$ )	45 min
Angioi et al. [50]	Dance	NR	Intra-rater test–retest (ICC 0.85–0.96) Inter-rater (ICC 0.89–0.97)	60 s
Chatfield [47]	Dance	Content Construct (discriminative, ANOVA)	Inter-rater ( $r = 0.95\text{--}0.96$ ) Intra-rater ( $r = 0.98$ )	4 min 4 s
Krasnow and Chatfield [44]	Dance	Content	Intra-rater ( $r = 0.82\text{--}0.95$ ) Inter-rater ( $r = 0.82\text{--}0.94$ )	NR
Barnett et al. [48]	Golf	Content Construct (discriminative, ANOVA)	Intra-rater (ICC 0.79) Test–retest (ICC 0.60)	NR
Lam and Zhang [43]	Racquetball	Content Criterion (concurrent $r = -0.48$ )	Intra-rater (ICC 0.87)	20–25 min
Barnett et al. [59]	Resistance training	Content Construct (regression model)	Inter-rater (ICC 0.67) Coefficient of variation (4.9 %)	8–10 min
Lubans et al. [45]	Resistance training	Content Construct (regression model)	Test–retest Rank order repeatability (ICC 0.67–0.87) Change in mean ( $r = -0.26$ to 0.21) Typical error (range 0.6–1.2; $r = -0.26$ to 0.21)	8–10 min
Myer et al. [42]	Resistance training	Content	NR	NR
Erbaugh [56]	Swimming	Content	Test–retest ( $r = 0.84\text{--}0.97$ ) Inter-rater ( $r = 0.98\text{--}0.99$ 85.9–96.4 %) Inter-trial ( $r = 0.97\text{--}0.99$ )	20 min
Zetou et al. [55]	Swimming	NR	Intra-rater ( $r = 0.81$ ) Inter-rater ( $r = 0.79$ )	NR
Buszard et al. [51]	Tennis	NR	Inter-rater (ICC 0.81–0.91)	NR
Farrow et al. [57]	Tennis	Ecological	Test–retest ( $r = 0.82$ forehand; $r = 0.80$ backhand)	NR
Messick [53]	Tennis	NR	Inter-rater (69–100 %) Intra-rater (96 %)	NR

ANOVA analysis of variance, ICC intraclass correlation coefficient, NR not reported

45, 47, 50, 52, 54, 56, 59]. Three tests took 5 min or less to assess a single participant [47, 50, 54]. Additionally, the resistance training skills battery reported 8- to 10-min test durations [45, 59]. The remaining three articles reported a test duration of 20 min or more [43, 52, 56]. The rest of the articles ( $n = 9$ ) included in this review either made no mention of the time needed to administer the given test, or

the time needed was unclear, thus test duration could not be determined.

### 3.5 Samples and Skills Tested

Information pertaining to skills tested and participant samples used can be found in Table 2. Overall, samples of



included studies were young, ranging from preschool age to college students, with the exception of two dance studies [47, 50] that included participants aged up to 30 years. Additionally, three of the 17 studies, all of which were dance tests [44, 47, 50], used some elite or professional dancers.

## 4 Discussion

This review was conducted to assess the methodological properties, validity, reliability, and test duration of process-oriented lifelong physical activity measurement tools, as well as to clearly define the characteristics unique to lifelong physical activities. Although 17 studies were included in this review, only assessments for eight different lifelong physical activities were identified (i.e., resistance training, badminton, tennis, cycling, racquetball, swimming, golf, and dance). All but one study reported some form of reliability, but fewer studies reported the validity of measurement tools. These results may indicate that, while some work has been done on creating valid tests of lifelong physical activities, current tests can still be improved. This review also highlighted the need for assessments of other popular lifelong physical activities, such as yoga, Pilates, tai chi, aerobics, and running.

### 4.1 Risk of Bias

It should be noted that the majority of the studies failed to describe the participants' characteristics in sufficient detail, which limits the generalizability of findings. For example, few studies described their sampling frame and participants' ability levels. While nine studies specifically stated the participants' skill levels (e.g., beginner, expert), one study [50] used all professional, national, or international level participants, and all five studies used all beginner level participants [43, 47, 51, 54, 57]. By using participants with high ability levels (e.g., professional), the applicability of the content tested for the general or even amateur population may be questionable. For example, competencies in rhythmical accuracy, spatial skills, and accuracy of movements may be too detailed for anyone other than the most elite dancers. Thus, while tests of dance competency exist [44, 47, 50], their suitability for assessing lifelong physical activity competency may be inadequate. In the future, recruiting a more heterogeneous sample with older people (above the age of 20 years) and varied ability levels, may be beneficial, as results may therefore be more applicable to the population as a whole. Thus, the validity and reliability of these lifelong physical activity assessments should hold true for people of all ages. If developed tests are not valid or reliable in older populations, then

identifying specific movement skill deficiencies in these populations may be compromised.

### 4.2 Reliability

As a whole, reliability was better reported than validity. Inter-rater reliability was the most commonly reported type of reliability. Three studies reporting inter-rater reliability had moderate reliability [49, 55, 59], two studies ranged from moderate to excellent levels of reliability [46, 53], and the rest of the studies reporting inter-rater reliability had excellent levels. Intra-rater reliability was also well reported, and levels of intra-rater reliability were classified as excellent for all these studies, except one study that was near excellent levels with an ICC value of 0.79 for a test of golf proficiency [48]. Rank order repeatability showed moderate to excellent levels of reliability for the resistance skill training battery, and acceptable levels of change in mean and typical error were also displayed for this test [45].

Test–retest reliability was only reported in four studies and should be a focus of future studies to see whether results are reliable over time, as opposed to a one-off measurement. If a test is to be considered reliable, the test needs to have adequate stability (i.e., results are similar over time) [39] and sensitivity (i.e., ability to detect small, meaningful differences in scores, such as in the resistance skills training battery) [45, 59, 60]. By addressing these issues, future tests can be administered with greater confidence regardless of time between assessments. While rank order repeatability is an important form of reliability, researchers are encouraged to assess other forms of test–retest reliability. More specifically, change in mean and typical error can be used to determine variability within an individual's score, which is particularly important when determining the effect of an intervention on movement skill competency.

### 4.3 Validity

Only ten of the studies included in this review reported validity. Overall, content validity was the most frequently cited type of validity, while criterion validity was largely unreported. Very few process-oriented measures of lifelong physical activities are available; thus, comparing results of one assessment with results of a second assessment for the same activity rarely occurs. Particular attention in future research should be given to ensuring additional forms of validity (e.g., predictive, construct), as opposed to just content validity, are established for any test of movement skill competency. Research is also required to create multiple assessments for a given sport or activity, thereby allowing for more construct and criterion validity of lifelong physical activities to be established. By creating more appropriate tests, researchers and practitioners alike will

possess a range of assessments to test an individual's competency, which can help to eliminate deficiencies in movement skills or better teach individuals how to correctly perform a skill. It is important to remember that test validity is highly contextual and is not carried across situations, thus it cannot be assumed that a test validated with children will provide similar results for adolescents or adults.

One reason that previous skill tests using process-oriented measures, such as the TGMD-2, have been used with success in the past may be due to the numerous types of validity that have been established in a number of different settings. For example, the content validity of the TGMD-2 was established through the agreement of three experts who judged the appropriateness of the skills included in the battery. Second, criterion validity was shown through the strong correlation of the TGMD-2 to a similar measure of movement ability. Finally, construct validity was established through its ability to test for age differentiation, group differentiation, item validity, subtest correlations (i.e., locomotor and object control subtest), and factor analysis [36].

Researchers are encouraged to assess the predictive validity of lifelong physical activity movement skill tests by comparing results with physical activity behavior. This is important because if lifelong physical activities are able to predict high levels of physical activity, then justification for the inclusion of such activities in the school curriculum, particularly in secondary school, may be warranted. Due to the decline in physical activity that commonly occurs in adolescence [61], it is imperative that young people develop competency in a range of fundamental, specialized, and lifelong physical activity movement skills. Indeed, recent reviews and national guidelines have highlighted the importance of developing movement skill competency to ensure that young people are prepared for a lifetime of physical activity [19, 62–64]. While the relationship between FMS and physical activity during childhood and adolescence has been well documented [4], less evidence is available to support the importance of FMS beyond the adolescent years. It is also well reported that not all individuals will attain proficiency in FMS. As such, these individuals may need an additional set of movement skills in lifelong physical activities that they can learn and may provide another or further opportunity to be physically active. Thus, lifelong physical activities may play a critical role in obtaining higher levels of physical activity into adulthood.

#### 4.4 Test Duration

Just under half of the studies in this review noted a test duration between 1 and 45 min. Longer tests may be

acceptable for smaller groups of people, while larger groups may be better served by a quick, efficient test for assessing skill competency. Unfortunately, there is no well-accepted criteria for determining whether a test is too short or too long; thus, researchers need to use their best judgment when creating tests [65]. Given that tests of lifelong physical activities may be targeted in schools, where lack of time in physical education is a known barrier [66], the need for shorter tests may be justified. Test duration may be influenced by other variables such as equipment needed, number of trials tested, and administration duties. While these are all important to consider when determining appropriate test duration, the validity and reliability of a given test should not be compromised. Previously, reviews have noted movement skill tests that take anywhere from 15 to 90 min to complete for a single participant [31, 32]. Around 20 min seems to be the most common amount of time used to assess various FMS, sport, and lifelong movement skills [31, 32]. For example, the TGMD-2 [36] takes about 20 min to administer, and this movement skill assessment is widely used [35, 67, 68]. More research on test duration for skill assessment may be beneficial to see approximately what amount of time balances feasibility with obtaining sufficient information on an individual's ability.

#### 4.5 Limitations

Limitations of this review are that only eight different lifelong physical activities were identified. More tests of lifelong physical activity competency may exist; however, either validity or reliability of these tests have not been established or they may appear elsewhere, but not in the peer-reviewed literature (e.g., yoga, Pilates). Another limitation is the lack of diverse samples tested. Few tests assessed non-elite and older aged individuals, thus applicability to the general population may be questioned. In addition, test–retest reliability was lacking, as this was only displayed in four studies. Thus, one-time measures of competency seem to be an issue in the assessment of lifelong physical activities.

#### 5 Conclusion

Lifelong physical activity movement skills may be advantageous for individuals to learn due to their individual or small group nature and as an opportunity to broaden their physical activity confidence and competence. Additionally, their need for little structure, decreased contact, varying levels of intensity and competitiveness, along with the ability to perform these activities into old age may allow individuals to be active at any age. A total of 17

studies were considered and reviewed for their methodological properties, validity, reliability, and test duration. Methodological characteristics, such as participants' details and stability of conditions need to be better reported in future studies. While moderate to excellent levels of intra-rater and inter-rater reliability were noted in the majority of tests, few tests of lifelong physical activities reported test-retest reliability. Validity was only reported in ten of the studies; content validity was the most common. Future research should look to establish additional forms of validity and reliability for current tests of lifelong physical activities. Tests of lifelong physical activity included in this review and created in the future should look to establish predictive validity in order to support the notion that competency in lifelong activities does allow for a lifetime of activity.

**Acknowledgments** The authors report no conflicts of interest within the information provided in this review. No funding was received by any of the authors to perform any portion of the review. Authorship criteria was met by all authors for this journal, and each author made a significant contribution to the final version of this paper.

## References

- Lloyd RS, Faigenbaum AD, Stone MH, et al. Position statement on youth resistance training: the 2014 international consensus. *Br J Sports Med.* 2014;48:498–505. doi:10.1136/bjsports-2013-092952.
- Stodden DF, Langendorfer SJ, Robertson MA. The association between motor skill competence and physical fitness in young adults. *Res Q Exerc Sport.* 2009;80(2):223–9.
- Stodden DF, True LK, Langendorfer SJ, et al. Associations among selected motor skills and health-related fitness: indirect evidence for Seefeldt's proficiency barrier in young adults? *Res Q Exerc Sport.* 2013;84(3):397–403.
- Barnett LM, van Beurden E, Morgan PJ, et al. Childhood motor skill proficiency as a predictor of adolescent physical activity. *J Adolesc Health.* 2009;44(3):252–9.
- Gallahue DL, Ozmun JC, Goodway J. Understanding motor development: infants, children, adolescents, adults. 7th ed. Boston: McGraw-Hill; 2012.
- Morgan PJ, Barnett LM, Cliff DP, et al. Fundamental movement skill interventions in youth: a systematic review and meta-analysis. *Pediatrics.* 2013;132(5):e1361–83. doi:10.1542/peds.2013-1167.
- Clark JE, Metcalfe JS. The mountain of motor development: a metaphor. In: Clarke JE, Humphrey JH, editors. *Motor development: research and reviews*, vol. 2. Reston: National Association for Sport and Physical Education; 2002. p. 163–90.
- Burton AW, Miller DE, Miller D. *Movement skill assessment*. Champaign: Human Kinetics Champaign; 1998.
- Hardy LL, Reinten-Reynolds T, Espinel P, et al. Prevalence and correlates of low fundamental movement skill competency in children. *Pediatrics.* 2012;130(2):e390–8. doi:10.1542/peds.2012-0345.
- Goodway JD, Robinson LE, Crowe H. Gender differences in fundamental motor skill development in disadvantaged preschoolers from two geographical regions. *Res Q Exerc Sport.* 2010;81(1):17–24.
- Seefeldt V. The concepts of readiness applied to motor skill acquisition. In: Magill RA, Ash MJ, Smoll FL, editors. *Children in sport*. Champaign: Human Kinetics; 1982. p. 31–7.
- Strong WB, Malina RM, Blimkie CJ, et al. Evidence based physical activity for school-age youth. *J Pediatr.* 2005;146(6):732–7.
- Coalter F. Sport and recreation in the United Kingdom: flow with the flow or buck the trends? *Manag Leis.* 1999;4(1):24–39.
- Fairclough S, Stratton G, Baldwin G. The contribution of secondary school physical education to lifetime physical activity. *Eur Phys Educ Rev.* 2002;8(1):69–84.
- Green K. Mission impossible? Reflecting upon the relationship between physical education, youth sport and lifelong participation. *Sport Educ Soc.* 2014;19(4):357–75. doi:10.1080/13573322.2012.683781.
- Green K. Lifelong participation, physical education and the work of Ken Roberts. *Sport Educ Soc.* 2002;7(2):167–82. doi:10.1080/1357332022000018850.
- Stodden DF, Goodway JD, Langendorfer SJ, et al. A developmental perspective on the role of motor skill competence in physical activity: an emergent relationship. *Quest.* 2008;60(2):290–306.
- Kirk D. Physical education, youth sport and lifelong participation: the importance of early learning experiences. *Eur Phys Educ Review.* 2005;11(3):239–55.
- Centers for Disease Control and Prevention. *Comprehensive school physical activity programs: a guide for schools*. Atlanta, GA: US Department of Health and Human Services; 2013.
- Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181–8.
- Dunn AL, Ross AE, Jakicic JM. Lifestyle physical activity interventions: history, short- and long-term effects, and recommendations. *Am J Sport Med.* 1998;15(4):398–412.
- Penney D, Jess M. Physical education and physically active lives: a lifelong approach to curriculum development. *Sport Educ Soc.* 2004;9(2):269–87.
- Ross JG, Dotson CO, Gilbert GG, et al. What are kids doing in school physical education? *J Phys Educ Recreat Dance.* 1985;56(1):73–6.
- Pangrazi R. *Dynamic physical education for elementary school children*. 15th ed. San Francisco: Benjamin Cummings; 2007.
- De Loe M. Epidemiology of sports injuries in the Swiss organization. *Int J Sports Med.* 1995;16(2):134–8.
- De Loe M, Goldie I. Incidence rate of injuries during sport activity and physical exercise in a rural Swedish municipality: incidence rates in 17 sports. *Int J Sports Med.* 1988;9(6):461–7.
- Nicholl J, Coleman P, Williams B. The epidemiology of sports and exercise related injury in the United Kingdom. *Br J Sports Med.* 1995;29(4):232–8.
- Australian Bureau of Statistics. *Participation in sport and physical recreation, Australia, 2011–2012*. 2013. <http://www.abs.gov.au/ausstats/abs@.nsf/Products/4177.0~2011-12~+Features~Characteristics+of+participation?OpenDocument>. Accessed 12 Nov 2014.
- Hands BP. How can we best measure fundamental movement skills? Paper presented at the Australian Council for Health, Physical Education and Recreation Inc. (ACHPER). In: 23rd biennial national/international conference: interactive health and physical education. 2002. Launceston, TAS.
- Nicholls R, Fleisig G, Elliott B, et al. Baseball: accuracy of qualitative analysis for assessment of skilled baseball pitching technique. *Sports Biomech.* 2003;2(2):213–26.
- Cools W, De Martelaer K, Samaey C, et al. Movement skill assessment of typically developing preschool children: a review

- of seven movement skill assessment tools. *J Sports Sci Med*. 2009;8(2):154.
32. Robertson SJ, Burnett AF, Cochrane J. Tests examining skill outcomes in sport: a systematic review of measurement properties and feasibility. *Sports Med*. 2014;44(4):501–18.
  33. Hardy LL, King L, Farrell L, et al. Fundamental movement skills among Australian preschool children. *J Sci Med Sport*. 2010;13(5):503–8.
  34. Pang AW-Y, Fong DT-P. Fundamental motor skill proficiency of Hong Kong children aged 6–9 years. *Res Sports Med*. 2009;17(3):125–44.
  35. Spessato BC, Gabbard C, Valentini N, et al. Gender differences in Brazilian children's fundamental movement skill performance. *Early Child Devel Care*. 2013;183(7):916–23.
  36. Ulrich DA. *Test of gross motor development*. 2nd ed. Austin: PRO-ED, Inc.; 2000.
  37. Sireci SG. The construct of content validity. *Soc Indic Res*. 1998;45(1/3):83–117.
  38. Barrow HM, McGee R, Tritschler KA. *Practical measurement in physical education and sport*. 4th ed. Philadelphia: Lea &Febiger; 1989.
  39. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008.
  40. Helmerhorst HJ, Brage S, Warren J, et al. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int J Behav Nutr Phys Act*. 2012;9(1):103–57.
  41. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–3.
  42. Myer GD, Kushner AM, Brent JL, et al. The back squat: a proposed assessment of functional deficits and technical factors that limit performance. *Strength Cond J*. 2014;36(6):4–27.
  43. Lam ET, Zhang JJ. The development and validation of a racquetball skills test battery for young adult beginners. *Meas Phys Educ Exerc Sci*. 2002;6(2):95–126.
  44. Krasnow D, Chatfield SJ. Development of the “performance competence evaluation measure”: assessing qualitative aspects of dance performance. *J Dance Med Sci*. 2009;13(4):101–7.
  45. Lubans DR, Smith JJ, Harries SK, et al. Development, test-retest reliability, and construct validity of the resistance training skills battery. *J Strength Cond Res*. 2014;28(5):1373–80.
  46. Ducheyne F, De Bourdeaudhuij I, Lenoir M, et al. Children's cycling skills: development of a test and determination of individual and environmental correlates. *Accid Anal Prev*. 2013;50:688–97.
  47. Chatfield SJ. A test for evaluating proficiency in dance. *J Dance Med Sci*. 2009;13(4):108–14.
  48. Barnett LM, Hardy LL, Brian A, et al. The development and validation of a golf swing and putt skill assessment for children. *J Sports Sci Med*. 2015;14(1):147–54.
  49. Toriola AL, Toriola OM, Dhaliwal HS, et al. Relationship between physical education students' achievements in a french badminton service test and expert ratings of technique quality. *Percept Mot Skills*. 2004;98(2):406–8.
  50. Angioi M, Metsios GS, Twitchett E, et al. Association between selected physical fitness parameters and aesthetic competence in contemporary dancers. *J Dance Med Sci*. 2009;13(4):115–23.
  51. Buszard T, Farrow D, Reid M, et al. Modifying equipment in early skill development: a tennis perspective. *Res Q Exerc Sport*. 2014;85(2):218–25.
  52. Macarthur C, Parkin PC, Sidky M, et al. Evaluation of a bicycle skills training program for young children: a randomized controlled trial. *Injury Prevention*. 1998;4(2):116–21.
  53. Messick JA. Prolongitudinal screening of hypothesized developmental sequences for the overhead tennis serve in experienced tennis players 9–19 years of age. *Res Q Exerc Sport*. 1991;62(3):249–56.
  54. Wang J, Liu W, Moffit J. Steps for arm and trunk actions of overhead forehand stroke used in badminton games across skill levels. *Percept Mot Skills*. 2009;109(1):177–86.
  55. Zetou E, Nikolaos V, Evaggelos B. The effect of instructional self-talk on performance and learning the backstroke of young swimmers and on the perceived functions of it. *J Phys Educ Sport*. 2014;14(1):27–35.
  56. Erbaugh SJ. Assessment of swimming performance of preschool children. *Percept Mot Skills*. 1978;46(3 Pt 2):1179–82.
  57. Farrow D, Reid M. The effect of equipment scaling on the skill acquisition of beginning tennis players. *J Sports Sci*. 2010;28(7):723–32.
  58. Lubans DR, Morgan P, Callister R, et al. Test–retest reliability of a battery of field-based health-related fitness measures for adolescents. *J Sports Sci*. 2011;29(7):685–93.
  59. Barnett L, Reynolds J, Faigenbaum AD, et al. Rater agreement of a test battery designed to assess adolescents' resistance training skill competency. *J Sci Med Sport*. 2015;18(1):72–6. doi:10.1016/j.jsams.2013.11.01253.
  60. Smith JJ, Morgan PJ, Plotnikoff RC, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. *Pediatrics*. 2014;134(3):e723–31.
  61. Dumith SC, Gigante DP, Domingues MR, et al. Physical activity change during adolescence: a systematic review and a pooled analysis. *Int J Epidemiol*. 2011;40(3):685–98.
  62. Hills AP, Dengel DR, Lubans DR. Supporting public health priorities: recommendations for physical education and physical activity promotion in schools. *Prog Cardiovasc Dis*. 2015;57(4):368–74. doi:10.1016/j.pcad.2014.09.010.
  63. Centers for Disease Control and Prevention. *Comprehensive school physical activity programs: a guide for schools*. 2013. <http://www.cdc.gov/healthyyouth/physicalactivity/cspap.htm>. Accessed 10 Nov 2014.
  64. Logan SW, Robinson LE, Wilson AE, et al. Getting the fundamentals of movement: a meta-analysis of the effectiveness of motor skill interventions in children. *Child Care Health Dev*. 2012;38(3):305–15.
  65. Streiner DL. *22 A checklist for evaluating the usefulness of rating scales. A guide for the statistically perplexed: selected readings for clinical researchers*. Toronto: University of Toronto Press; 2013. pp. 267–288.
  66. Morgan PJ, Hansen V. Classroom teachers' perceptions of the impact of barriers to teaching physical education on the quality of physical education programs. *Res Q Exerc Sport*. 2008;79(4):506–16.
  67. Bryant ES, Duncan MJ, Birch SL. Fundamental movement skills and weight status in British primary school children. *Eur J Sport Sci*. 2014;14(7):730–6. doi:10.1080/17461391.2013.870232.
  68. Hardy LL, Barnett L, Espinel P, et al. Thirteen-year trends in child and adolescent fundamental movement skills: 1997–2010. *Med Sci Sports Exerc*. 2013;45(10):1965–70.