

Original article

# Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age

Patricia E. Longmuir<sup>a,b,\*</sup>, Charles Boyer<sup>a</sup>, Meghann Lloyd<sup>c</sup>, Michael M. Borghese<sup>a,d</sup>,  
Emily Knight<sup>a</sup>, Travis J. Saunders<sup>a,e</sup>, Elena Boiarskaia<sup>f</sup>, Weimo Zhu<sup>f</sup>, Mark S. Tremblay<sup>a,b</sup>

<sup>a</sup> Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario K1H 8L1, Canada

<sup>b</sup> Department of Pediatrics, Faculty of Medicine and School of Human Kinetics, University of Ottawa, Ottawa K1H 8M5, Canada

<sup>c</sup> Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa L1H 7K4, Canada

<sup>d</sup> School of Kinesiology and Health Studies, Queen's University, Kingston, Ontario K7L 3N6, Canada

<sup>e</sup> Department of Applied Human Sciences, University of Prince Edward Island, Charlottetown, Prince Edward Island C1A 4P3, Canada

<sup>f</sup> Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received 14 May 2015; revised 3 July 2015; accepted 24 July 2015

Available online 11 November 2015

## Abstract

**Purpose:** The primary aim of this study was to develop an assessment of the fundamental, combined, and complex movement skills required to support childhood physical literacy. The secondary aim was to establish the feasibility, objectivity, and reliability evidence for the assessment.

**Methods:** An expert advisory group recommended a course format for the assessment that would require children to complete a series of dynamic movement skills. Criterion-referenced skill performance and completion time were the recommended forms of evaluation. Children, 8–12 years of age, self-reported their age and gender and then completed the study assessments while attending local schools or day camps. Face validity was previously established through a Delphi expert ( $n = 19$ , 21% female) review process. Convergent validity was evaluated by age and gender associations with assessment performance. Inter- and intra-rater ( $n = 53$ , 34% female) objectivity and test–retest ( $n = 60$ , 47% female) reliability were assessed through repeated test administration.

**Results:** Median total score was 21 of 28 points (range 5–28). Median completion time was 17 s. Total scores were feasible for all 995 children who self-reported age and gender. Total score did not differ between inside and outside environments (95% confidence interval (CI) of difference:  $-0.7$  to  $0.6$ ;  $p = 0.91$ ) or with/without footwear (95%CI of difference:  $-2.5$  to  $1.9$ ;  $p = 0.77$ ). Older age ( $p < 0.001$ ,  $\eta^2 = 0.15$ ) and male gender ( $p < 0.001$ ,  $\eta^2 = 0.02$ ) were associated with a higher total score. Inter-rater objectivity evidence was excellent (intraclass correlation coefficient (ICC) = 0.99) for completion time and substantial for skill score (ICC = 0.69) for 104 attempts by 53 children (34% female). Intra-rater objectivity was moderate (ICC = 0.52) for skill score and excellent for completion time (ICC = 0.99). Reliability was excellent for completion time over a short (2–4 days; ICC = 0.84) or long (8–14 days; ICC = 0.82) interval. Skill score reliability was moderate (ICC = 0.46) over a short interval, and substantial (ICC = 0.74) over a long interval.

**Conclusion:** The Canadian Agility and Movement Skill Assessment is a feasible measure of selected fundamental, complex and combined movement skills, which are an important building block for childhood physical literacy. Moderate-to-excellent objectivity was demonstrated for children 8–12 years of age. Test–retest reliability has been established over an interval of at least 1 week. The time and skill scores can be accurately estimated by 1 trained examiner.

© 2017 Production and hosting by Elsevier B.V. on behalf of Shanghai University of Sport. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Agility course; Children; Dynamic motor skill; Locomotor skill; Object manipulation; Population assessment

## 1. Introduction

Physically active lifestyles are important to children's health, in both the short and long term.<sup>1–3</sup> Physical literacy is the motivation,

confidence, physical competence, knowledge, and understanding to value and take responsibility for engagement in physical activities for life.<sup>4</sup> The ability to monitor progress along the lifelong physical literacy journey requires a broad spectrum of valid and reliable methods for charting progress. For children 8–12 years of age, existing measures enable monitoring of motivation and confidence<sup>5,6</sup> and many aspects of physical fitness.<sup>7–9</sup> However, the physical competence domain of physical literacy includes not only

Peer review under responsibility of Shanghai University of Sport.

\* Corresponding author.

E-mail address: [plongmuir@cheo.on.ca](mailto:plongmuir@cheo.on.ca) (P.E. Longmuir)

fitness, but also physical movement competence. Movement skill, health-related physical fitness, and the child's self-perceived motor skill competence combine to influence the child's physical activity.<sup>10</sup> Among young children, the relationship between movement skill and physical activity is weak.<sup>10</sup> Among older children and adolescents, the strength of the movement skill–physical activity relationship becomes increasingly important.<sup>10</sup> While younger children will persist with physical activity regardless of their movement competence,<sup>10</sup> older children who lack competence in fundamental movement skills, such as running, jumping, balancing, kicking, catching, and throwing, are less likely to adopt an active lifestyle.<sup>11–13</sup> Thus, movement competence is an increasingly important determinant of physical literacy as children mature.

In describing the modern understanding of physical literacy, Whitehead<sup>14</sup> categorized movement capabilities as fundamental, combined, or complex. Simple movement capabilities, such as core stability, balance, coordination, or speed variation,<sup>15</sup> are combined to create more advanced movement capabilities. For example, balance, core stability, and movement control are combined to provide equilibrium.<sup>15</sup> These combined movement capabilities enable the development of complex movements, such as bilateral, inter-limb and hand–eye coordination, control of acceleration/deceleration, and turning, twisting, and rhythmic movements.<sup>15</sup> Giblin et al.<sup>15</sup> have discussed the limitations of current assessments in relation to the broader concept of physical literacy. Current assessments of movement skill require that children perform each skill in isolation,<sup>16–19</sup> making them time and resource intensive to administer.<sup>15,20</sup> The static testing environment and performance of isolated skills do not assess combined and complex movement capabilities<sup>21</sup> or reflect the open, dynamic, and complex physical activity environments typical of childhood play.<sup>22</sup>

The purpose of this study was to develop an assessment of children's movement capabilities that would incorporate not only fundamental movement skill execution, but evidence of the child's ability to combine simple movements and demonstrate complex movement capabilities. Children with higher levels of physical literacy will be able to control body movements, speed, and balance in order to optimize their performance for each physical activity environment. The goal was an assessment that would be feasible for population surveillance, as well as monitoring movement skills over time. The secondary aim of this study was to establish the validity, objectivity, and reliability evidence for the newly-developed movement skill assessment.

## 2. Methods

The protocol for this study was approved by the Research Ethics Board of the Children's Hospital of Eastern Ontario Research Institute, and the research approval boards/committees of all collaborating school boards and day camps. Before the children attempted the course, parents completed a screening questionnaire to determine whether their child was known to have a medical condition that might be affected by completion of the assessment activities. The study was conducted in 3 phases, from 2007 to 2012. Initially, expert consultations were combined with an environmental scan of physical education curricula and published research to develop the activities included in the initial

assessment (2007–2009). Feasibility of the assessment was established through an iterative design process with students in Grades 4–6 (2009–2011). In the final phase (2011–2012), cross-sectional assessments were used to generate evidence of the validity and objectivity of the Canadian Agility and Movement Skill Assessment (CAMSA). Children performed the CAMSA during 1 study visit, while each performance was video recorded. Video recordings were used to establish the intra-rater and inter-rater objectivity of the assessment. Repeated performance of the assessment across intervals of 3–14 days evaluated test–retest reliability.

### 2.1. CAMSA development

The initial aim of this study was to develop an assessment that would evaluate children's movement skills as well as their ability to combine simple movement capabilities and perform more complex movement skills in response to a changing environment. A sense of one's own physical capabilities combined with adept interactions with one's environment is a foundational concept for physical literacy.<sup>14</sup> An international expert advisory panel comprised of physical educators and physical activity scientists selected a dynamic series of movement skills as the preferred format. A combined measure of skill execution and completion time was recommended because it was expected that children who are more advanced in their physical literacy journey would select the appropriate balance of speed and skill for optimal performance.<sup>23</sup> The advisory panel also felt that the agility course format would be enjoyable for the children and would also enable the relatively quick assessment of groups of children as required for population surveillance. Existing measures were not felt to be appropriate because they either assessed fundamental movement skills in isolation<sup>16,17,19</sup> or examined agility without regard for movement skill.<sup>24</sup> The use of equipment and space that would typically be available in a school or recreation setting was also an important criterion during task selection. The initial assessment design required jumps on 2 feet, 1-footed hopping, throwing and catching on the run, dodging from side to side, and the kick of a soccer ball. This combination of skills had established performance criteria<sup>19</sup> and was intended to represent the movement patterns that children would be expected to develop through school physical education classes. The advisory panel chose not to include complex movements involving twisting, climbing and rotation in order to minimize the risk of injury among children with lower levels of physical literacy.

An iterative process of design and evaluation was followed to establish the feasibility of the CAMSA. Feasibility assessments were completed by 596 children ( $n = 274$ , 46% female) in Grades 4, 5, and 6 (aged 8–12 years). The space required for the initial assessment design was found to be too large for many elementary school gymnasias. There was also substantial variability between different appraisers in the scoring of the dodging skill, whereby children had to move from side to side between a series of cones (unpublished data). In response, a smaller course was designed that required children to traverse the course twice in order to complete all of the required skills.

The dodging activity was replaced by the side slide, which had established reference criteria.<sup>19</sup>

The revised CAMSA required children to travel a total distance of 20 m while completing 7 movement skill tasks (Fig. 1). The movement skills selected were: (1) 2-footed jumping into and out of 3 hoops on the ground, (2) sliding from side to side over a 3 m distance, (3) catching a ball and then (4) throwing the ball at a wall target 5 m away, (5) skipping for 5 m, (6) 1-footed hopping in and out of 6 hoops on the ground, and (7) kicking a soccer ball between 2 cones placed 5 m away. As recommended by the international Delphi panel ( $n = 19$ ),<sup>23</sup> children who were able to balance the speed and skill components of the assessment obtained the highest raw score. The order of the tasks was chosen to optimize the safety of participating children and require multiple transitions between different types of skills. To reduce the probability that a bouncing ball might pose a hazard to participants, the soccer ball kick was the final skill performed and a soft ball, which would not bounce, was used for the throwing and catching task. Hopping and jumping tasks were separated by sliding, running, and skipping movements to increase the balance and complexity of the skill transitions required.

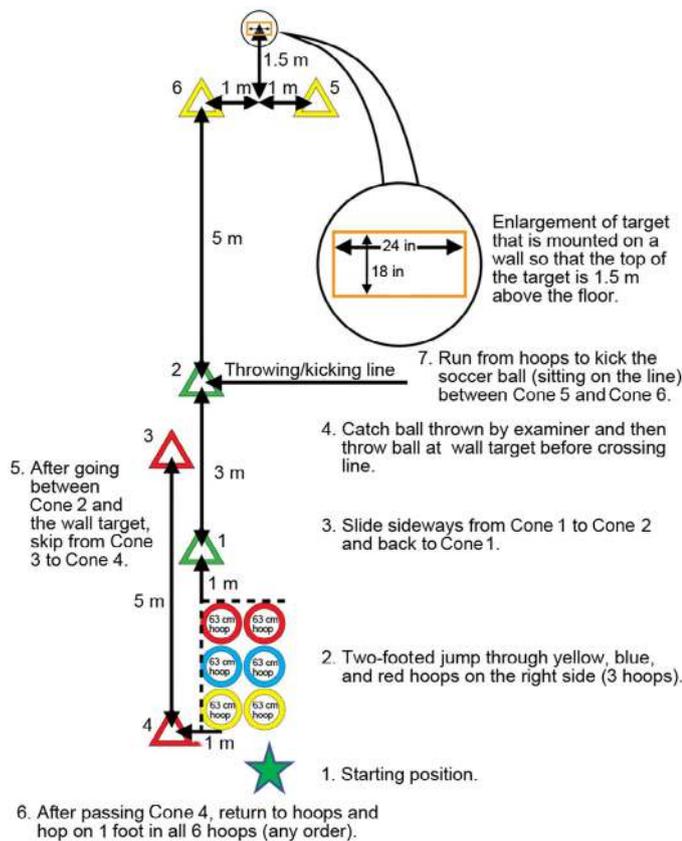


Fig. 1. CAMSA course schematic. Children start at the hoops and work their way through the cones in numbered order before going through the hoops a second time. A video demonstration of the CAMSA is available on the website of the Canadian Assessment of Physical Literacy (<https://www.capl-ecsf.ca/capl-training-videos>). CAMSA = Canadian Agility and Movement Skill Assessment.

### 2.1.1. Administration of the CAMSA

Groups of children were instructed to complete the assessment as fast as possible while performing the skills to the best of their ability. The Delphi expert panel achieved consensus that measuring both speed and movement skill was important. The rationale was that a child with greater physical literacy would be able to select the correct speed for optimal skill performance, while those with lower physical literacy would go very slowly in order to perform the required skills or very fast without regard for skill quality.<sup>23</sup> The assessment was demonstrated twice for each group of children. The first demonstration was done slowly, with each skill explained as it was demonstrated. The second demonstration indicated the effort and speed required. Each child performed 2 timed and scored trials. Timing started on the “go” command and ended when the participant kicked the soccer ball. Verbal cues were given throughout the assessment, prior to each skill, to minimize the impact of memory on task sequence and completion time. Verbal cues were used only to remind the participant of the next task to be performed. No feedback was provided on task performance and no attempt was made to encourage or alter the child’s performance.

Two examiners were required to administer and score the assessment. The first examiner measures the completion time, throws the ball for the catching task, places the ball for the kicking task and provides the verbal cues regarding the next task to be performed. The second examiner evaluates the quality of each skill performed according to established criteria (Table 1).

Table 1  
Canadian Agility and Movement Skill Assessment skill and time scoring criteria<sup>a</sup>.

Skill criteria <sup>b</sup>	Time criteria	
	Completion time (s)	Score
1. Jump with 2 feet in and out of yellow, blue, and red hoops	<14.0	14
2. No extra jumps and no touching the hoops when 2-footed jumping	14.0–14.9	13
3. Body and feet are aligned sideways sliding in 1 direction	15.0–15.9	12
4. Body and feet aligned sideways sliding in opposite direction	16.0–16.9	11
5. Touch cone when changing directions after sliding	17.0–17.9	10
6. Catches ball (no drop or trap against body)	18.0–18.9	9
7. Uses overhand throw to hit target	19.0–19.9	8
8. Transfers weight and rotates body when throwing	20.0–20.9	7
9. Correct step-hop foot pattern when skipping	21.0–21.9	6
10. Alternates arms and legs when skipping, arms swinging for balance	22.0–23.9	5
11. Land in each hoop when hopping on 1 foot	24.0–25.9	4
12. Hops only once in each hoop (no touching of hoops)	26.0–27.9	3
13. Smooth approach to kick ball between cones	28.0–29.9	2
14. Elongated stride on last stride before kick impact with ball	≥30.0	1

<sup>a</sup> Total raw score = skill score + time score (range 1–28 points).

<sup>b</sup> One point is awarded for each skill criterion performed correctly. If the skill is performed incorrectly or only partially performed (e.g., smooth approach to kick ball but ball does not go between cones for criterion #13), 0 point is awarded. Maximum skill score is 14 points.

Examiners during all phases of this research were research assistants and graduate students with post-secondary degrees in kinesiology. All examiners completed at least 3 h of additional training specific to this protocol. The training included detailed reviews of the movement quality criteria as well as repeated practice trials for both the timing and skill examiner.

The time required to complete the course was recorded, and then converted to a point score (range 1–14 points; [Table 1](#)) based on the range of completion times among all children in our study (range 11–36 s; median 17 s). The quality of each skill (2-footed jump, side slide, catch, throw, skipping, 1-footed hop, and kick) was scored as either performed (score of 1) or not observed (score of 0) across 14 reference criteria (0–14 points). Skill criteria were drawn from the Test of Gross Motor Development Version 2.<sup>19</sup> Product-based criteria (i.e., scoring based on the outcome of the skill attempt) were the primary form of movement quality assessment. The expert advisory group felt that a focus on performance outcomes would enable the assessment to be more relevant to the child's ability to successfully participate in active play with peers. It was also desirable that children whose quality of movement was affected by a disability ( $n = 25$ ) would not automatically be given a lower score. The total score (maximum 28 points) was calculated as the sum of the skill and time scores. Details of the scoring criteria are described in [Table 1](#).

### 2.1.2. Development of a 14-point motor skill score

The CAMSA initially utilized a 20-point scoring system for skill assessment. The 20-point score was based on established criteria for the performance of each skill<sup>19</sup> but required extensive training to implement in a reliable manner. To enhance the feasibility of the assessment scoring procedures, a Rasch model was fitted to the initial 20-point score. A Rasch analysis compares the data collected to what would be expected based on a theoretical model.<sup>25</sup> Standard convention then assigned each item to a difficulty rating: easy ( $< -0.7$  logits), medium ( $-0.7$  to  $0.7$  logits), or hard ( $> 0.7$  logits). Items with similar logit scores and low uniqueness were removed or combined in order to maintain a similar range of item difficulty within a 14-point skill score that was expected to be easier to administer. A Spearman rank order correlation found that the 14- and 20-point skill scores for children were highly correlated ( $n = 729$ ,  $r = 0.95$ ,  $p < 0.001$ ). Therefore, the 14-point scoring system was used for these analyses ([Table 1](#)).

## 2.2. Study participants

### 2.2.1. Participant recruitment

To establish the psychometric integrity of the CAMSA, a convenience sample of children, 8–12 years of age, was recruited for this study via local schools and day camps. Two school boards in the Ottawa area invited principals to have their schools participate in the study. If a school responded to the invitation, individual classroom teachers at that school were offered the opportunity to take part in the study, and their students were provided with a study information package. Day camps for children were offered through a local university. The

day camp programs utilized the sport and recreation facilities of the university. Most camps offered a variety of activities (e.g., recreational activities, games, sports) but some were designed to teach a specific sport (e.g., tennis camp). Single sport day camps designed for competitive athletes were not eligible for the study. Camp managers agreed that researchers could wait at the morning check-in area to invite children, accompanied by their parent/guardian, into the study. For both school and day camp recruitment, parent's written consent and child's verbal assent were obtained prior to the child being enrolled in the study. All children whose parents provided written consent and who agreed to complete the study activities were included in the study. Children with disabilities affecting skill performance were able to participate to the best of their ability. Final scores were calculated only for those children who performed all of the assessment activities without modification. Children whose parents indicated, via a health screening questionnaire, that they should not participate in vigorous activity due to health concerns ( $n = 9$ ) were excluded from all sessions.

### 2.2.2. Description of participants

In order to characterize the participants, children were asked to self-report their age and gender using a questionnaire. No attempt was made to establish the biological age or sex of each child. Canadian Health Measures Survey protocols<sup>26</sup> were used to assess participant's height to the nearest 0.1 cm using a portable stadiometer (seca GmbH, Hamburg, Germany), and weight to the nearest 0.1 kg with a digital scale (A&D Medical, Milpitas, CA, USA), while children were wearing light clothing and no footwear. In order to assess whether our sample was representative of the Canadian population of children of this age, normative data from the Centers for Disease Control and Prevention were used to determine children's age- and sex-adjusted body mass index (BMI) percentile scores.<sup>27</sup>

### 2.3. CAMSA validity

A Delphi panel,<sup>28</sup> comprised of those with backgrounds in childhood physical activity and motor skill development, was used to establish the face validity of the skill components and the weighting of the completion time and skill scores.<sup>23</sup> The Delphi process gathered expert opinions through 3 rounds of guided questions. Round 1 utilized open-ended questions to enable the participants to share their expertise related to movement skill assessment in children. In Round 2, experts rated a series of statements using a 5-point Likert scale. The statements were generated by summarizing areas of agreement and divergent opinion among the Round 1 responses. In Round 3, experts had an opportunity to re-evaluate their previous Likert scale ratings in light of a summary of the Round 2 responses. The Delphi process was designed to develop consensus recommendations without undue influence from peer comments<sup>28</sup> where limited empirical data are available.<sup>29</sup> It has previously been used to validate physical activity<sup>30</sup> and motor-function<sup>31</sup> assessments. The Delphi panel for this project reached consensus on combining the skill and time scores, enabling an evaluation of the impact of self-reported age and gender on total score,

completion time and skill performance. Age and gender associations were examined to evaluate convergent validity.

To facilitate use of the CAMSA in a broad range of settings, including countries with limited facilities, the skill score and completion time were evaluated to determine the impact of: (a) footwear compared to bare feet, and (b) indoor compared to outdoor settings. The comparison of footwear to bare feet was completed in a gymnasium, with the order of footwear condition randomized among the participants. Results from the gymnasium with footwear were compared to performance of the CAMSA on an outdoor grass field, again with the order of indoor/outdoor condition randomized among participants. Each repetition of the assessment was administered by the same examiners, regardless of condition.

#### 2.4. CAMSA objectivity and reliability

Test-retest reliability was assessed by having children perform the CAMSA on 2 separate days. The interval between test dates ranged from 2 to 14 days, depending on the school or camp schedule. Retest intervals were categorized as short (2–4 days) or long (8–14 days) for analysis. In order to assess inter- and intra-rater objectivity for timing and skill score, children were video recorded (with consent) while performing the assessment. Performances were scored by 7 examiners on 3 separate occasions, with an interval of 2–3 days between viewing sessions. Completion time was calculated twice by 4 examiners over an interval of 4–5 days; however, 1 examiner completed the second viewing session after only a 1-day interval due to time constraints. Excluding this individual did not affect the results, thus results are presented using data from all 4 examiners. The order in which the video performances were reviewed was randomized between examiners. The technique of using video-recorded performances to establish inter/intra-rater objectivity of physical activity and fitness has previously been reported.<sup>7,32</sup> This method was selected as the method of choice for this study because children would not be able to exactly replicate their performance on multiple trials. The use of video-recordings enabled the examiners to evaluate the exact same performance on multiple days. It also enabled the same performance to be scored from the same viewing angle by multiple examiners, which would not have been possible given the gymnasium testing environment.

#### 2.5. Statistical analyses

Descriptive data are described as mean  $\pm$  SD or median (range), as appropriate. All analyses were performed in SPSS Version 22.0 (IBM Corp., Armonk, NY, USA). Level of statistical significance was set at  $p < 0.05$ . Self-reported age and gender were mandatory variables in all regression models.

##### 2.5.1. Evaluation of potential sample bias

An evaluation of the associations between the number of practice trials completed and CAMSA total score, skill score and completion time was completed. The aim of this analysis was to enhance assessment feasibility by minimizing administration time (i.e., practice trials required) while avoiding a significant learning effect on performance. Children performed 1

( $n = 205$ , 63% female), 2 ( $n = 39$ , 46% female), or 3 ( $n = 450$ , 57% female) practice trials of the entire assessment before being assessed during 2 additional trials. A one-way univariate ANCOVA indicated that, after adjusting for self-reported age (scores significantly increased with age,  $p < 0.001$ ) and gender (higher scores for boys,  $p < 0.001$ ), the number of practice trials did not significantly influence the best total score ( $p = 0.36$ ). The number of practice trials was not associated with the 14-point skill score ( $p = 0.45$ ) or the completion time score ( $p = 0.08$ ) from the best scored trial. Since number of practice trials did not influence the assessment, all subsequent analyses were performed without regard to practice trial completion.

Since day camp recruitment was primarily from camps designed to teach specific sports (e.g., learn to play tennis or soccer), the source of participants (camp vs. school) was also evaluated relative to total score, skill score and completion time as an assessment of potential sample bias. A one-way univariate ANCOVA found no association between the recruitment source for participants (camp vs. school) and the total CAMSA score ( $p = 0.18$ ), skill score ( $p = 0.94$ ) or completion time score ( $p = 0.19$ ). Therefore data from all participants were analyzed together without regard to recruitment site.

##### 2.5.2. Analyses of CAMSA validity

Details of the analyses for the Delphi process have previously been reported.<sup>23</sup> Briefly, open-ended responses from Round 1 were qualitatively analyzed to identify areas of commonality and diverse opinion. Each area was then represented by 1 or more statements that were evaluated using a 5-point Likert scale during Rounds 2 and 3. Tabulation of the Likert scale responses indicated consensus, defined as at least 75% agreement. Analyses of variance evaluated total score, completion time and skill score relative to expected patterns based on age (better scores with older age) and gender (males score higher than females). Percentile scores were calculated to inform the interpretation of assessment results. Paired  $t$  tests examined total score under the footwear versus bare feet and indoor versus outdoor conditions.

##### 2.5.3. Analyses of CAMSA objectivity and reliability

Intraclass correlation coefficient (ICC), two-way random single measures for absolute agreement (ICC 2,1), reported with 95% confidence interval (CI), were used to evaluate the level of agreement across trials to establish evidence for inter-rater and intra-rater objectivity and test-retest reliability. The strength of agreement was determined by the following criteria for the ICC value: 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, excellent.<sup>33</sup> For results with less than substantial objectivity or reliability, a paired  $t$  test was used to determine if the difference between tests was significant. To determine the number of occasions or examiners needed to obtain acceptable objectivity and reliability, the collected data were further analyzed using GENOVA (Version 3.1; Robert L. Brennan, University of Iowa, Iowa City, IA, USA), a software for generalizability theory.<sup>34</sup> Generalizability theory is an extension of reliability modeling that can estimate the reliability of a measure that has multiple sources of potential error.<sup>35</sup>

### 3. Results

#### 3.1. Description of participants

In total, 1165 children (598 (51%) females), 8–12 years of age, participated across all phases of this study. Recruitment rates were 76% for school testing in 2011/2012 and 77% for schools tested in 2012/2013. Recruitment rates could only be calculated for the school-based assessments because total camper registration numbers were not available. The 91 children who did not self-report their age (30 (33%) females) were excluded from the analyses. These 91 children did not differ from those who provided their age for either self-reported gender ( $t = 1.39, p = 0.17$ ) or testing site (school vs. camp;  $t = 1.8, p = 0.08$ ).

The characteristics of the participants for each phase of the study are provided in Table 2. Data for age and gender were based on the child's self-report. Height ( $n = 838, 477$  (57%) females) and weight ( $n = 822, 464$  (56%) females) collected during the 2011/2012 and 2012/2013 school-based assessments indicated that study participants were  $144.9 \pm 9.5$  cm in height (range 121.5–177.5 cm; mean height percentile  $68.4\% \pm 31.0\%$ ) and  $39.7\% \pm 11.1$  kg in weight (range 22.3–97.7 kg; mean weight percentile  $65.7\% \pm 30.2\%$ ) in weight. Their mean BMI percentile score was  $61.6\% \pm 28.7\%$  (range <1 to >99), with 27.9% classified as overweight or obese. The distribution of self-reported gender among study participants differed by recruitment location (school vs. camp;  $t = 5.7, p < 0.001$ ). Boys and girls were relatively equally distributed when recruitment was done through local schools (56% female). Children attending day camps were primarily male (65%).

#### 3.2. Feasibility evidence

The panel of 19 internationally recognized experts in childhood movement skill, fitness and physical literacy assessment recommended that the same scoring system be used regardless of child's self-reported age or gender.<sup>23</sup> The completion time and skill scores were to be equally weighted. Interpretation of the score should differ by age and gender such that assessment feedback would be relative to expected performance by age and gender.

Table 2  
Self-reported age (mean  $\pm$  SD) and gender (%) of study participants by study phase.

Study analysis	Female	Male	Total
<b>Convergent validity and percentile score calculation</b>			
<i>n</i> (%)	526 (53)	469 (47)	995 (100)
Age (year)	10.0 $\pm$ 1.2	10.1 $\pm$ 1.4	10.1 $\pm$ 1.3
<b>Test–retest reliability evidence</b>			
<i>n</i> (%)	28 (47)	32 (53)	60 (100)
Age (year)	9.6 $\pm$ 1.1	10.1 $\pm$ 1.2	9.9 $\pm$ 1.2
<b>Inter- and intra-rater video objectivity</b>			
<i>n</i> (%)	18 (34)	35 (66)	53 (100)
Age (year)	10.1 $\pm$ 1.6	9.9 $\pm$ 1.6	10.0 $\pm$ 1.6
<b>Indoor vs. outdoor</b>			
<i>n</i> (%)	13 (46)	15 (54)	28 (100)
Age (year)	10.4 $\pm$ 1.3	10.6 $\pm$ 1.3	10.5 $\pm$ 1.3
<b>Barefoot vs. footwear</b>			
<i>n</i> (%)	13 (45)	16 (55)	29 (100)
Age (year)	10.1 $\pm$ 1.4	10.9 $\pm$ 1.0	10.6 $\pm$ 1.3

Table 3

Percentiles for total Canadian Agility and Movement Skill Assessment score by self-reported age and gender.

	Age (year)				
	8	9	10	11	12
Number of boys ( $n = 469$ )	20	108	179	117	45
5th Percentile score	10.0	13.0	13.7	18.0	19.0
25th Percentile score	14.0	17.0	19.3	22.0	22.0
50th Percentile score	17.0	20.0	22.0	24.0	24.0
75th Percentile score	19.0	23.0	24.0	26.0	26.0
95th Percentile score	24.0	26.0	26.0	27.0	27.0
Number of girls ( $n = 526$ )	20	113	235	127	31
5th Percentile score	11.0	13.0	12.0	16.0	16.8
25th Percentile score	13.3	17.0	18.0	20.0	21.0
50th Percentile score	17.5	19.0	20.0	22.0	23.0
75th Percentile score	20.5	21.8	23.0	24.0	25.0
95th Percentile score	23.0	24.0	25.0	26.7	27.0

Note: Range for the theoretically possible total score was 1–28 points.

Total CAMSA scores could be calculated, using the 14-point criterion scale, for all 995 study participants who self-reported age and gender. Median time score was 11 points (maximum 14, range 1–14) and the median skill score was 12 points (range 2–14). Median total score (time + skill) was 22 points (range 3–28). Percentile scores for total assessment score were calculated by age and gender (Table 3). For each age and gender category, the 5th, 25th, 50th, 75th, and 95th percentiles for total score were determined to assist with the interpretation of assessment results.

Total CAMSA score was calculated for 2 timed and scored trials among 840 (472 (56%) females) children. Almost all of these children were recruited through local schools ( $n = 788$ ). Total score from Trial 1 (median 21 of 28 points; range 6–28) was significantly lower (95%CI of difference: 0.16 to 0.48;  $p < 0.001$ ) than total score for Trial 2 (median 22 points; range 5–28). Time scores improved significantly on the second trial (mean difference:  $0.4 \pm 1.3$  points, 95%CI of difference: 0.28 to 0.46;  $p < 0.001$ ). There was no difference in skill score between the 2 trials (95%CI of difference:  $-0.07$  to 0.17;  $p = 0.42$ ). Participants with 2 timed and scored trials were more likely to be female ( $\chi^2 = 23.9, p < 0.001$ ) and tested in school rather than summer camp ( $\chi^2 = 414.9, p < 0.001$ ), but were similar in age ( $t = 1.2, p = 0.11$ ) to children with only 1 timed and scored trial. Trial 1 was missing for 3 children tested in schools (2 females). Trial 2 was missing for 189 children (96 (51%) females, 4 (2%) gender unknown, 152 (81%) from day camps), primarily due to time constraints imposed on the data collection session by the host school/camp.

A paired-samples  $t$  test indicated that there was no difference in total CAMSA score between indoor and outdoor conditions (95%CI of difference:  $-0.7$  to 0.6;  $p = 0.91$ ). Total score did not differ significantly when, indoors, the assessment was performed with and without footwear (95%CI of difference:  $-2.5$  to 1.9;  $p = 0.77$ ). The time required to set up the CAMSA, typically 5–7 min, is less than or similar to the time required to set up the equipment for tests of motor development.

#### 3.3. Validity evidence

In a multivariable model ( $p < 0.001, \eta^2 = 0.17, R^2 = 0.16$ ), older age ( $p < 0.001, \eta^2 = 0.15$ ) and male gender ( $p < 0.001,$

$\eta^2 = 0.02$ ) were significantly associated with a higher total assessment score (maximum 28 points). Estimated marginal means  $\pm$  SE were  $21.8 \pm 0.2$  for boys and  $20.9 \pm 0.2$  for girls. Estimated marginal mean scores increased from  $17.8 \pm 0.6$  for 8 years old to  $23.9 \pm 0.4$  for 12 years old. The same pattern of associations was observed for the skill ( $p < 0.001$ ,  $\eta^2 = 0.04$ ,  $R^2 = 0.04$ ) and time ( $p < 0.001$ ,  $\eta^2 = 0.20$ ,  $R^2 = 0.20$ ) scores.

### 3.4. Objectivity and reliability evidence

Objectivity was assessed across 104 attempts by 53 children (34% female). Evidence for inter-rater objectivity was substantial (ICC = 0.69) for the skill score and excellent for the completion time (ICC = 0.997) (Table 4). Evidence for intra-rater objectivity (Table 4) for the skill score was moderate (ICC = 0.52) and excellent for completion time (ICC = 0.996) (Table 4). Scores on the second trial (median 18 of 28 points, range 10–20) were significantly higher ( $p < 0.001$ ) than on the first trial (median 17 points, range 9–20), with the mean difference being 0.57 (95%CI: 0.35 to 0.78). The intra-rater objectivity for the time score did not change with the removal of the examiner who had only a 1-day interval

Table 4  
Intra- and inter-rater objectivity evidence for skill score and completion time<sup>a</sup>.

Type of test	ICC	95%CI	Strength of ICC <sup>e</sup>
<b>Skill score</b>			
<i>Inter-rater<sup>b</sup></i>			
All trials	0.69	0.61, 0.76	Substantial
Trial 1	0.70	0.61, 0.79	Substantial
Trial 2	0.66	0.55, 0.77	Substantial
<i>Intra-rater<sup>c</sup></i>			
All examiners	0.52	0.43, 0.60	Moderate
Examiner 1	0.45	0.20, 0.64	Moderate
Examiner 2	0.55	0.33, 0.72	Moderate
Examiner 3	0.43	0.19, 0.63	Moderate
Examiner 4	0.52	0.28, 0.69	Moderate
Examiner 5	0.49	0.26, 0.67	Moderate
Examiner 6	0.57	0.35, 0.73	Moderate
Examiner 7	0.53	0.30, 0.70	Moderate
<b>Completion time</b>			
<i>Inter-rater<sup>b</sup></i>			
All trials	0.997	0.995, 0.998	Excellent
Trial 1	0.997	0.994, 0.998	Excellent
Trial 2	0.993	0.990, 0.995	Excellent
<i>Intra-rater<sup>d</sup></i>			
All examiners	0.996	0.995, 0.997	Excellent
Examiner 1	0.999	0.999, 1.000	Excellent
Examiner 2	0.998	0.998, 0.999	Excellent
Examiner 3	0.991	0.986, 0.994	Excellent
Examiner 4	0.996	0.994, 0.997	Excellent

<sup>a</sup> Data for these analyses were from 53 children (18 (34%) females) who performed a total of 104 trials of the Canadian Agility and Movement Skill Assessment.

<sup>b</sup> ICC calculated across 7 examiners for each trial. All 7 examiners repeated the evaluation for 3 trials at an interval of 2–3 days between trials.

<sup>c</sup> ICC calculated separately for each examiner across 2 trials/examiners at an interval of 2–3 days between trials.

<sup>d</sup> ICC calculated across 4 examiners for each trial. Interval between ratings was 1–4 days.

<sup>e</sup> Conventions for the strength of objectivity assessments: moderate: 0.41–0.60; substantial: 0.61–0.80; excellent: 0.81–1.00.<sup>33</sup>

Abbreviations: ICC = interclass correlation coefficient; 95%CI = 95% confidence interval.

Table 5  
Decision study of Generalizability Theory results for participants  $\times$  occasion  $\times$  examiner.

Examiner	G-coefficient		
	Occa-1	Occa-2	Occa-3
<b>CAMSA skill score</b>			
1	0.84271	0.91464	0.94143
2	0.91464	0.95542	0.96983
3	0.94143	0.96983	0.97968
4	0.95542	0.97720	0.98468
5	0.96401	0.98168	0.98771
6	0.96983	0.98468	0.98974
7	0.97403	0.98684	0.99119
<b>CAMSA completion time</b>			
1	0.99487	0.99743	—
2	0.99743	0.99871	—
3	0.99828	0.99914	—
4	0.99871	0.99936	—

Abbreviation: CAMSA = Canadian Agility and Movement Skill Assessment.

between sessions. The results of the decision (D study) on skill scores and completion time using Generalizability Theory are summarized in Table 5. Using a G-coefficient of 0.8 as the cut-off score, it was demonstrated that the participants’ performance on the CAMSA can be reliably estimated using 1 trained examiner’s 1-time rating.

Evidence for test–retest reliability for completion time was excellent across short ( $n = 59$ ; ICC = 0.84; 95%CI: 0.74 to 0.91) and long ( $n = 16$ ; ICC = 0.82; 95%CI: 0.53 to 0.93) test intervals. Evidence for test–retest reliability for the skill score was moderate ( $n = 44$ ; ICC = 0.46; 95%CI: 0.20 to 0.66) over a short interval but substantial ( $n = 16$ ; ICC = 0.74; 95%CI: 0.42 to 0.90) over a long interval. Although reliability evidence was moderate for the movement skill score measured over a short interval, there was no significant difference between the test (median 17 of 28 points; range 11–20) and retest (median 17; range 8–20) scores ( $n = 59$ ; 95%CI of difference:  $-0.2$  to  $0.9$ ;  $p = 0.22$ ).

## 4. Discussion

### 4.1. Measuring movement skill in a dynamic environment

To enhance their physical literacy journey, children need to be competent in a variety of movement skills.<sup>14,36</sup> They also require the ability to perform those skills in response to changing environments<sup>36</sup> and to combine and coordinate fundamental skills to create more complex movement patterns.<sup>21</sup> This study developed an objective assessment that incorporates selected fundamental (jump, slide, catch, throw, skip, hop kick), complex (hand–eye coordination, control of acceleration/deceleration, rhythmic movement, inter-limb coordination) and combined (balance, core stability, coordination, equilibrium, precision) movement skills<sup>15</sup> within an agility and movement skill course. Although the skills and movements required by this course were selected to represent children’s movement capacity from 8 to 12 years of age, it must be recognized that many other aspects of agility and movement skill (e.g., bilateral coordination, twisting, dexterity, climbing) are not assessed by this protocol. The Delphi

panel supported the choice of movements included in this protocol as being reflective of the movement skills that children of this age should acquire through school physical education classes.<sup>23</sup>

Median completion time per trial was 17 s (range 11.2–41.4 s). Although the number of practice trials did not significantly influence total scores, 2 practice trials followed by 2 measured trials are recommended to ensure that children can give a good effort on both measured trials. The resulting assessment time of 1.5–2.0 min per child is comparable to the time required for fitness protocols currently used for population surveillance.<sup>26</sup>

There are 2 approaches to addressing the assessment of movement skill as children grow and mature. One approach administers different tasks and interprets scores based on age groupings, with the assessment procedures increasing in complexity for older children (e.g., Movement ABC-2<sup>37</sup>). The alternative approach utilizes the same tasks across all ages but varies the performance criteria or scoring to reflect differences in expected performance with age (e.g., Test of Gross Motor Development-2<sup>19</sup>). As recommended through a Delphi expert review process,<sup>23</sup> this project used the same assessment for all children (8–12 years of age). This approach would enable the tracking of movement skill over time. Raw score interpretation based on age- and gender-matched percentile scores ensured that performance feedback was relative to the expected performance for age- and gender-matched peers.

#### 4.2. CAMSA feasibility

The final version of the CAMSA was successfully performed by all 1165 children (598 (51%) females). There was a small proportion of children ( $n = 91$ ) whose total score could not be interpreted relative to percentile scores by age and gender due to missing data for age. Age was self-reported by children in the current study, and the researchers did not otherwise have access to demographic information. Comparisons of outdoor versus indoor terrain and wearing shoes compared to bare feet suggested that the assessment could be suitable for international research, including developing countries where facilities and footwear may not be readily available.

Lengthy assessment times and low examiner: participant ratios have previously been identified as the limiting factors that inhibit population-level surveillance of movement skill. For example, the ALPHA study of children's physical fitness (20 m shuttle run, handgrip, long jump, anthropometric measures) required approximately 2.5 h to complete an assessment of 20 children,<sup>38</sup> a time the authors estimated would double with the inclusion of a movement skill assessment.<sup>20,39</sup> In this study, mean completion time was 17 s for each repetition of the assessment. Thus overall assessment time was estimated to be 25 min for 20 children (examiner demonstration = 1 min; 17 s/practice trial  $\times$  2 trials/child  $\times$  20 children = 12 min; 17 s/measured trial  $\times$  2 trials/child  $\times$  20 children = 12 min). The much shorter time required to set up (5–7 min) and administer (25 min for 20 children) our assessment of movement skill increases the feasibility of monitoring movement skill development at a population level. It also increases the feasibility of comprehensive physical literacy assessments, which include not only

movement skill but also measures of physical fitness, motivation for activity, activity knowledge, and daily behavior.<sup>40</sup>

All of the examiners for this study had extensive experience in movement skill analysis. They had graduate degrees in kinesiology and had completed up to 5 h of additional training specific to the protocol used in this study. The assessment protocol currently requires 2 examiners in order to obtain valid scores. In schools or other settings, this may require a physical education and classroom teacher to jointly administer the assessment. Alternatively, 2 teachers could combine their classes in order to administer the assessment. One examiner assesses skill quality while a second examiner times the performance, provides the cues for each skill and enables the catch, throw and kicking skills. Future research should evaluate the feasibility and reliability of the assessment protocol when it is administered by examiners with less movement skill analysis experience. Future research should also examine the training required for the role of the second examiner (timing, *etc.*) to evaluate the validity and reliability of the assessment by 1 skilled examiner plus an assistant.

#### 4.3. CAMSA validity

This research examined the convergent validity of the CAMSA results. Face validity of the CAMSA had previously been established.<sup>23</sup> Among our study participants, self-reported age and gender were associated with CAMSA performance. Older children obtained higher skill scores and completed the CAMSA faster than younger children. These results are similar to previous reports that have linked improvements in running speed, agility, and balance with increasing age among children 6–11 years of age.<sup>41</sup>

In this study, raw performance scores were higher for boys compared to girls among the older (9–12 years old) but not younger (8 years old) participants. An evaluation of individual skill components suggested that these gender differences were primarily due to boys obtaining higher scores for the ball (throwing, catching, and kicking) skills. These results are consistent with previous research and recognized developmental patterns.<sup>13,41–43</sup> For example, among 276 slightly older children (mean age = 10.1 years, 146 (53%) females), males performed better than females in kicking, overhand throwing and catching with no gender difference during assessments of vertical jump, hopping and side gallop. Girls typically excel at hopping, skipping and small muscle coordination.<sup>41</sup> The small effect size for gender observed for the total score appears to reflect the combination of skills chosen for the assessment. Girls would be expected to excel at some tasks (e.g., hopping, skipping) while boys would be expected to excel at others (e.g., kicking, throwing). Alternatively, another potential explanation of the small gender effect is the relatively young age of the study participants. Most of the children in this study were 9, 10, or 11 years of age. A larger gender effect may have been observed with a larger sample of older (>11 years) children.

It is important to note that the age effect or small gender effect observed would not be expected to impact the motivation of assessment participants. The age- and gender effects were evident only for raw scores, which are not divulged to the

children completing the course. The Delphi panel recommended that only age- and gender-adjusted scores be shared with children and parents because the same tasks are performed across a broad age range where maturity and movement skill would be expected to differ.<sup>23</sup>

#### 4.4. CAMSA objectivity and reliability

This study examined intra- and inter-rater objectivity and test–retest reliability of the CAMSA. All objectivity and reliability measures of the completion time were excellent. Although inter-rater objectivity for the skill score was substantial and intra-rater skill score objectivity was moderate, a decision study using Generalizability Theory found that the skill score could be reliably estimated based on a 1-time rating by a trained examiner. These results suggest that assessment administration time could potentially be further reduced by requiring only 1 timed and scored performance instead of the 2 used for this study. However, replication of these results with a larger group of children and with a larger proportion of females is needed before such a change could be recommended.

Test–retest skill score reliability was lower over shorter (2–4 days) compared to longer intervals (8–14 days). Children may remember the CAMSA over a short interval, and are therefore able to enhance their performance via a learning effect when multiple trials are performed over a short interval. Such a learning effect was not detected over a longer test–retest interval, suggesting the assessment is suitable for monitoring performance over time provided the interval is more than 4 days. A minimum test–retest interval of 1 week is recommended to minimize the influence of previous assessment attempts.

#### 4.5. Strengths and limitations

The strengths of this study include the large sample size, the balanced participation of both males and females, and the broad age range among study participants. The possibility of recruitment bias must always be considered in studies of physical activity (those who are more skilled are more likely to participate). We recruited 938 of 1165 participants from summer camps. The 1165 participants enrolled in the study were 77% of all children approached. Of children recruited from schools, 27.9% were overweight or obese according to their measured height and weight, a proportion virtually identical to data for the Canadian population (28% of children considered to be overweight or obese<sup>44</sup>). These results suggest that the study participants were representative of the population of Canadian children and fit and athletic children were not over-represented in our study.

It is important to recognize that any assessment of agility and movement skill must select from among the almost endless array of fundamental, complex and combined skills that can be performed in a diverse range of environments. The fundamental, complex and combined skills included in the CAMSA were felt to be representative of the skills that should be developed by children of this age through school physical education classes.<sup>23</sup> Evidence of validity for the CAMSA was difficult to determine due to the lack of an existing “gold standard” protocol. Age- and gender patterns among the assessment results reflected what is known

from other measures (skills increase with age; boys excel at some skills while girls excel at others). Direct comparisons to existing measures of motor skill would be difficult as the established protocols require the performance of individual skills in isolation.<sup>17,19,45</sup> In addition, existing measurement protocols are not suitable to assess movement skill throughout childhood. Some existing protocols (e.g., Test of Gross Motor Development-2<sup>19</sup>) are limited to a specific age range. Other protocols (e.g., Movement ABC-2<sup>37</sup>) change the assessment tasks at different ages.

The inter-rater and intra-rater objectivity testing was done solely with children attending day camps. Due to scheduling commitments, it was not possible to administer repeat assessments to the school-based participants. The study population that was recruited from the day camps may have been more active and skilled than the general population, as all of these children had chosen to attend a sport camp. Camps designated for highly skilled athletes were excluded from the study recruitment, so that the included camps were specifically designed for children of all abilities. Nevertheless, the possibility that children would be more likely to register for a sport camp if they believed that they would enjoy the camp activity and be successful must be considered in the interpretation of the objectivity results.

## 5. Conclusion

The CAMSA is a feasible assessment of a sample of fundamental, complex, and combined movement skills among children 8–12 years of age. Assessment results suggest that the CAMSA accurately reflects known developmental changes in movement skill, and that the time and skill scores can be accurately estimated by 1 trained examiner. Test–retest reliability has been established over an interval of at least 1 week. The CAMSA offers an alternative approach to assess movement proficiency that is suitable for population surveillance or the assessment of groups of children in a relatively short period of time.

## Acknowledgments

The authors wish to thank Katie McClelland and Joel Barnes for their assistance with data collection and cleaning. This study was partially funded by a grant from the Canadian Institutes of Health Research awarded to Dr. Meghann Lloyd and Dr. Mark Tremblay (IHD 94356).

## Authors' contributions

PEL designed the study, collected data, analyzed data and drafted the manuscript; CB collected and analyzed data, and helped to draft the manuscript; ML conceptually designed the CAMSA and collected data; MMB collected and analyzed data; EK collected data; TJS collected data; EB analyzed data; WZ analyzed data; MST conceptually designed the CAMSA. All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

## Competing interests

The authors declare that they have no competing interests.

## References

- Centers for Disease Control and Prevention. *Physical activity and health: the benefits of physical activity*. Available at: <http://www.cdc.gov/physicalactivity/everyone/health/index.html>; 2008 [accessed 09.11.2008].
- Malina R. *Benefits of activity from a lifetime perspective*. Champaign, IL: Human Kinetics; 1994.
- Public Health Agency of Canada. *The benefits of physical activity: for children/youth*. Available at: <http://www.phac-aspc.gc.ca/pau-uap/fitness/benefits.html#1>; 2005 [accessed 09.11.2008].
- International Physical Literacy Association. *Definition of physical literacy*. Available at: <https://www.physical-literacy.org.uk/>; 2014 [accessed 06.01.2015].
- Hay JA. Adequacy in and predilection for physical activity in children. *Clin J Sport Med* 1992;**2**:192–201.
- Harter S. The perceived competence scale for children. *Child Dev* 1982;**53**:87–97.
- Boyer C, Tremblay MS, Saunders TJ, McFarlane A, Borghese M, Lloyd M, et al. Feasibility, validity and reliability of the plank isometric hold as a field-based assessment of torso muscular endurance for children 8 to 12 years of age. *Pediatr Exerc Sci* 2013;**25**:407–22.
- Tremblay MS, Shields M, Laviolette M, Craig CL, Janssen I, Connor Gorber S. Fitness of Canadian children and youth: results from the 2007–2009 Canadian Health Measures Survey. *Health Rep* 2010;**21**:7–20.
- Scott SN, Thompson DL, Coe DP. The ability of the PACER to elicit peak exercise response in the youth. *Med Sci Sports Exerc* 2013;**45**:1139–43.
- Stodden DF, Goodway JD, Langendorfer SJ, Robertson MA, Rudisill ME, Garcia C, et al. A developmental perspective on the role of motor skill competence in physical activity: an emergent relationship. *Quest* 2008;**60**:290–306.
- Bouffard M, Watkinson EJ, Thompson LP, Causgrove Dunn JL, Romanow SK. A test of the activity deficit hypothesis with children with movement difficulties. *Adapt Phys Act Q* 1996;**13**:61–73.
- Lloyd M, Saunders TJ, Bremer E, Tremblay MS. Long-term importance of fundamental motor skills: a 20-year follow-up study. *Adapt Phys Act Q* 2014;**31**:67–78.
- Lubans DR, Morgan PJ, Cliff DP, Barnett LM, Okely AD. Fundamental movement skills in children and adolescents: review of associated health benefits. *Sports Med* 2010;**40**:1019–35.
- Whitehead M. *Physical literacy throughout the lifecourse*. London: Routledge Taylor & Francis Group; 2010.
- Giblin S, Collins D, Button C. Physical literacy: importance, assessment and future directions. *Sports Med* 2014;**44**:1177–84.
- Bruininks RH. *Bruininks-Oseretsky test of motor proficiency*. Circle Pines, MN: American Guidance Service; 1978.
- Henderson SE, Sugden DA, Barnett AL, Smits-Engelsman CM. *Movement Assessment Battery for Children-2*. San Antonio, TX: Pearson; 2010.
- Folio MR, Fewell RR. *Peabody Development Motor Scales (PDMS-2)*. San Antonio, TX: Therapy Skill Builders; 2000.
- Ulrich DA. *Test of Gross Motor Development (TGMD-2)*. Austin, TX: PRO-ED; 2000.
- Wiat L, Darrah J. Review of four tests of gross motor development. *Dev Med Child Neurol* 2001;**43**:279–85.
- Tremblay MS, Lloyd M. Physical literacy measurement—the missing piece. *Phys Health Educ J* 2010;**76**:26–30.
- Watkinson EJ, Causgrove Dunn J. Applying ecological task analysis to the assessment of playground skills. In: Steadward R, Wheeler G, Watkinson EJ, editors. *Adapted physical activity*. Edmonton: University of Alberta Press; 2004.
- Francis CE, Longmuir PE, Boyer C, Andersen LB, Barnes JD, Boiarskaia E, et al. The Canadian Assessment of Physical Literacy: development of a model of children's capacity for a healthy, active lifestyle through a Delphi process. *J Phys Act Health* 2016;**13**:1–43.
- Sheppard JM, Young WB. Agility literature review: classifications, training and testing. *J Sports Sci* 2006;**24**:919–32.
- Zhu W, Cole EL. Many-faceted Rasch calibration of a gross motor instrument. *Res Q Exerc Sport* 1996;**67**:24–34.
- Tremblay MS, Langlois R, Bryan S, Esliger D, Patterson J. Canadian health measures survey pre-test: design, methods and results. *Health Rep* 2007;**18**:21–30.
- Kuczmariski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC growth charts for the United States: methods and development. *Vital Health Stat 11* 2002;(246):1–190.
- Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007;**12**:1–8.
- Gupta UG, Clarke RE. Theory and applications of the Delphi technique: a bibliography (1875–1994). *Technol Forecast Soc Change* 1996;**53**:185–211.
- Marques AI, Santos L, Soares P, Santos R, Oliveira-Tavares A, Mota J, et al. A proposed adaptation of the European Foundation for Quality Management Excellence Model to physical activity programmes for the elderly: development of a quality self-assessment tool using a Delphi process. *Int J Behav Nutr Phys Act* 2011;**8**:104. doi: 10.1186/1479-5868-8-104
- Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;**39**:214–23.
- Angioi M, Metsios GS, Twitchett E, Koutedakis Y, Wyon M. Association between selected physical fitness parameters and esthetic competence in contemporary dancers. *J Dance Med Sci* 2009;**13**:115–23.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- Brennan RL. *Generalizability theory*. New York, NY: Springer-Verlag; 2001.
- Morrow Jr. JR. Generalizability theory. In: Safrit MJ, Wood TM, editors. *Measurement concepts in physical education and exercise science*. Champaign, IL: Human Kinetics; 1998,p.73–96.
- PHE Canada. *What is physical literacy?* Available at: <http://www.phecanada.ca/programs/physical-literacy/what-physical-literacy>; 2014 [accessed 22.04.2014].
- Brown T, Lalor A. The movement assessment battery for children (2nd ed.): a review and critique. *Phys Occup Ther Pediatr* 2009;**29**:86–103.
- Williams HG, Pfeiffer KA, Dowda M, Jeter C, Jones S, Pate RR. A field-based testing protocol for assessing gross motor skills in preschool children: the children's activity and movement in preschool study motor skills protocol. *Meas Phys Educ Exerc Sci* 2009;**13**:151–65.
- Cools W, Martelaer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. *J Sports Sci Med* 2009;**8**:154–68.
- Longmuir PE. Understanding the physical literacy journey of children: the Canadian Assessment of Physical Literacy. *ICSSPE BULLETIN-J Sport Sci Phys Educ* 2013;**65**:276–82.
- Duger T, Bumin G, Uyanik M, Aki E, Kayihan H. The assessment of Bruininks-Oseretsky test of motor proficiency in children. *Pediatr Rehabil* 1999;**3**:125–31.
- Barnett LM, van Beurden E, Morgan PJ, Brooks LO, Beard JR. Gender differences in motor skill proficiency from childhood to adolescence: a longitudinal study. *Res Q Exerc Sport* 2010;**81**:162–70.
- Bar-Or O, Rowland TW. *Pediatric exercise medicine: from physiologic principles to healthcare application*. 2nd ed. Champaign, IL: Human Kinetics; 2004.
- Shields M, Tremblay MS. Canadian childhood obesity estimates based on WHO, IOTF and CDC cut-points. *Int J Pediatr Obes* 2010;**5**:265–73.
- Bruininks RH, Bruininks BD. *Bruininks-Oseretsky test of motor proficiency*. 2nd ed. San Antonio, TX: Pearson; 2005.