The Physical Educator Vol. 74 • pp. 441–475 • 2017
https://doi.org/10.18666/TPE-2017-V74-I3-7459

PEDAGOGY

Early Validation Evidence of a Canadian Practitioner-Based Assessment of Physical Literacy in Physical Education: Passport for Life

Ken R. Lodewyk and James L. Mandigo

Abstract

Physical and Health Education Canada has developed and implemented a formative, criterion-referenced, and practitioner-based national (Canadian) online educational assessment and support resource called Passport for Life (PFL). It was developed to support the awareness and advancement of physical literacy among PE students and teachers. PFL consists of three assessments for each of the four components (active participation, fitness, movement, and living skills). The aim of this study was to uncover initial validation evidence for its current uses using four of the five broad guidelines (content, response processes, internal structure, relations with other variables) for establishing satisfactory score validity as established by the Standards for Educational and Psychological Testing. We conducted a pilot test with 860 students in Grades 4 and 5 in 2013–2014. We analyzed these data, along with data collected across 2 years for Grades 3 to 6 (n = 1,036 in 2013–2014 and 1,254 in 2014–2015) and Grades 7 to 9 (n = 1,793 in 2013–2014 and 1,151 in 2014–2015). A portion (15 to 25%) of these students completed some of the assessments a second time in each of these years. Validation evidence included the development of PFL by a number of domain experts, the nature and format of the components and scales relative to existing literature and evidence, the administra-

Ken R. Lodewyk is an associate professor, Department of Kinesiology, Faculty of Applied Health Sciences, Brock University. James L. Mandigo is professor and vice-provost, Department of Kinesiology, Brock University. Please send author correspondence to klodewyk@brocku.ca

tive procedures to guide teachers to implement the assessments, the alignment of items with each component construct and scale, teacher feedback, and positive and significant relations and temporal (predictive) consistency over the 2 years within and across scales and components. These results provide general support for the PFL and its intended use and highlight several cautions and recommendations.

According to Physical and Health Education Canada (PHE Canada, n.d.), physical literacy is moving "with competence and confidence in a wide variety of physical activities in multiple environments that benefit the healthy development of the whole person" (para. 2). This definition has been supported in key policy documents such as the Ontario Health and Physical Education Curriculum (Ontario Ministry of Education, 2015), Canadian Sport Policy 2.0 (Sport Canada, 2012), and SHAPE America's (2014) National Standards for K–12 Physical Education. In recent years, PHE Canada has led the development of a national educational support resource called Passport for Life (PFL) to support the awareness, assessment, and advancement of physical literacy among physical education (PE) students and teachers (see http://passportforlife.ca/). PHE Canada designed PFL to serve as a feasible (i.e., clear, achievable, economical, efficient, safe, timely) formative, criterion-referenced, practitioner-based assessment of key aspects of physical literacy. The primary aim of this study was to uncover initial validation evidence for current uses of PFL, that is, to determine "the degree to which evidence and theory support the interpretation of test scores" for its proposed use (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014, p. 11). We also endeavored to explore the feasibility of PFL, namely, its usability for practitioners to administer in a timely manner at the risk of the results having somewhat diminished psychometric properties (Gosling, Rentfrow, & Swann, 2003).

In the position paper that helped to frame PHE Canada's definition of physical literacy, Mandigo, Francis, Lodewyk, and Lopez (2009) outlined the holistic nature of physical literacy within educational settings as a lifelong personal journey. Building on the work of Whitehead (2001), they integrated key concepts beyond the physical characteristics of physical literacy by ensuring that the concept was

consistent with current thinking around literacy (United Nations Educational, Scientific, and Cultural Organization, 2003). Concepts such as movement capacities that reflect the embodied self, physically challenging environments that motivate individuals across a number of activities, and incorporating key cognitive and social components by being able to "read" the environment were embedded into PHE Canada's definition based upon Whitehead's seminal work. The holistic nature of the physical literacy definition is consistent with the goals of lifelong psychomotor, health-related physical activity participation and fitness, and affective and cognitive learning outcomes that are embedded throughout PE curricula across Canada (Kilborn, Lorusso, & Francis, 2016) and elsewhere (Lacy & Hastad, 2003; SHAPE America, 2014). Because of the complexity and multidimensionality of such a holistic view of physical literacy, PFL was designed to reflect students' physical literacy closely (rather than exactly equate with or mirror it), so there is no composite score on it beyond scores for fitness skills, movement skills, active participation, and living skills. These four representative components of physical literacy in PFL were identified during a consultative process lasting several years and align with PHE Canada's definition of physical literacy. After completing a short demographic survey, participants proceed through the movement and fitness skill components. These include six physical movement assessments (three for each component), which the teacher administers during PE class. During these assessments, the teacher closely follows a detailed rubric to rate each student on a 4-point scale. Participants also complete the living skills (consisting of feeling, thinking, and relating scales) and active participation (comprising the scales called movement environment, and interest and intentions) components consisting of self-report items that students can complete online during or outside of school hours.

The consultation and development process of PFL also resulted in the emergence of several keystone guiding principles. These included the need for PFL to be inclusive, adaptable, individualized, relevant (e.g., aligned to curriculum), process (vs. product) oriented, safe, comprehensive, ongoing, affirming, valid, efficient and clear, and useful for formative (learning) rather than summative (grades) and the need for the promotion of lifelong development of physical

literacy. To support the use of PFL as a learning tool, PHE Canada collects and stores student data from PFL in an online platform, enabling teachers to input, manage, and download the data for each of their students and classes. This also allows teachers to use the data to provide individualized attention to help students and classes set goals and work toward continued improvement. The online PFL program also includes guides with information, resources, and ideas to help students, teachers, and parents interpret and use the assessment results for ongoing improvements in physical literacy. Students receive a paper Passport that includes a summary of their results on each completed assessment, along with suggestions for how to improve each component, to bring home to share with their parents/guardians.

The broad guidelines for establishing satisfactory score validity set by the Standards for Educational and Psychological Testing (Standards; AERA, APA, & NCME, 2014) served as the theoretical framework for this study. These standards assert that, rather than validating an instrument, validation involves providing ongoing evidence about the property of test scores and interpretations stemming from particular contextual uses of that instrument. Consequentially, an instrument can be neither invalid nor valid but rather relatively valid depending on its application. Further, instead of articulating validity as a variety of "types" (e.g., content, criterion, and construct), the Standards also promote a unified conceptualization of validity evidence stemming mainly from the content, response processes, internal structure, relations with other variables, and consequences of testing. Because we wish to describe the development and initial results and procedures of PFL, we will report on the first three with some early evidence for the fourth (relations with other variables). More specifically, for the first (score validity evidence that is based on content) we will analyze the format, wording, and administrative guidelines for the items and scales of the instrument. For the second (internal structure evidence), we will attest to how well the instrument-specific items align (interrelate) with its construct(s) and score interpretations. For the third (response process evidence), we will investigate the extent that scores reflect the construct, partic-

ularly through the way that participants engage with (i.e., interpret and answer) the items. Finally, we will explore relations with other variables to ascertain how closely the test scores and their interrelations associate with similar or associated variables concurrently or predictively as measured by other instruments.

Method

Participants

Pilot test. PFL was first developed for and pilot tested in the 2012–2013 academic year to 860 students in Grades 4 (n = 325; 37.8%) and 5 (n = 532; 61.6%) (and n = 3 not reporting) with relative equal balance between males (n = 355) and females (n = 351) from 18 schools in eight provinces (British Columbia, Alberta, Saskatchewan, Manitoba, Quebec, New Brunswick, Nova Scotia, and Newfoundland) of Canada.

Grades 3 to 6. Table 1 provides the sample sizes for each administration of PFL by component. In the fall season of the 2013–2014 academic year, 1,036 students in Grades 3 to 6 participated in PFL with approximately 40% of students completing all of the PFL components and approximately 22% (n = 123–240) completing some of the components a second time in the spring season. Grade level was reported by 90% of the students ($n_{\text{Grade 3}}$ = 109, $n_{\text{Grade 4}}$ = 221, $n_{\text{Grade 5}}$ = 218, $n_{\text{Grade 6}}$ = 244) and only 15% (n = 176) reported their gender (81 females, 95 males), so we did not perform any comparisons by gender for this report. Nine provinces were represented with the most from Ontario (n = 420) and the least from Prince Edward Island (n = 6).

The sample size of the 2014–2015 administration of PFL to Grades 3 to 6 was 1,254, which was reduced to 1,199 following a screening of outliers. Those providing voluntary demographic data revealed relatively equal proportions of participants by grade level ($n_{\text{Grade 3}}$ = 457, $n_{\text{Grade 4}}$ = 270, $n_{\text{Grade 5}}$ = 254, $n_{\text{Grade 6}}$ = 218) and gender (145 females, 182 males) with most from the province of Ontario (n = 573).

Table 1
Sample Sizes for the Passport for Life Components by Developmental Level, Year, and Time

			Assessment component and time (T)							
			Movement skills		Fitness skills		Living skills		Active participation	
Grade	Year	Provinces	T1	T2	T1	T2	T1	T2	T1	T2
3–6	2013–14	BC, AB, SK, MN, ON, QB, NB, PE, NL	647	200	791	240	892	123	1141	154
	2014–15	BC, AB, SK, MN, ON, QB, PE, NL, NT	629	151	817	187	980	220	1082	262
7–9	2013–14	BC, AB, MN, ON, NB, NL	1241	580	1349	559	1348	536	1549	547
	2014–15	BC, AB, SK, MN, ON, QB, NB	589	138	757	194	878	195	996	209

Note. BC = British Columbia; AB = Alberta ; SK = Saskatchewan; MN = Manitoba; ON = Ontario; QB = Quebec ; NB = New Brunswick; NS = Nova Scotia; PE = Prince Edward Island; NL = Newfoundland; NT = Northwest Territories.

Grades 7 to 9. In 2013–2014, 1,793 students in Grades 7 to 9 completed PFL. The sample was from six provinces with the majority from Ontario (n = 1,276) and the least from New Brunswick (n = 20) and Newfoundland (n = 22). Only a portion of the total sample reported their gender (301 females, 262 males) and grade level ($n_{\text{Grade 7}}$ = 53, $n_{\text{Grade 8}}$ = 72, $n_{\text{Grade 9}}$ = 431).

In 2014–2015, 1,151 students in Grades 7 to 9 completed PFL. Screening for outliers resulted in the deletion of 69 cases, so the final participant sample for analysis was 1,082. Participants were from seven of the Canadian provinces and territories with most (n = 616) from Ontario. Only a portion of the total sample (n = 598; 53%) reported their gender (316 females, 282 males). The number of students by grade was 443 in Grade 7, 273 in Grade 8, and 366 in Grade 9.

Data Collection

When participating teachers registered their class for the PFL program, they were provided with the option to input the data from the various measures directly into an online database. This database is housed on a secure (password and username protected) online repository hosted by PHE Canada that generates a Passport specific to each student and class. The data provide formative feedback to the teacher, students, and parents on current levels of each student's physical literacy. The data contain no individual identifiers. Each participant receives a unique code, which ensures anonymity. As such, it falls under the category of secondary data analysis, which is defined as ". . . the use in research of information or human biological material collected for a purpose other than the current research purpose" (Canadian Institutes of Health, Natural Science, and Engineering Research Council of Canada & Social Sciences and Humanities Research Council of Canada, 2014, p. 209). The university research ethics board and PHE Canada approved the use of these data for the purposes of the validation study outlined in this paper.

Measures

Feedback from practicing teachers. Following the completion of the initial PFL for each developmental level, most of the participating teachers ($n_{\text{Grade 4–5 pilot}}$ = 22, 100% participation; $n_{\text{Grade 7–9}}$ = 31, 68% participation) voluntarily completed an online exit survey devel-

oped by PHE Canada program management. The survey consisted of 55 items, for which participants responded on a 4-point Likert scale (1 = *strongly agree*, 4 = *strongly disagree*) and through written responses to provide feedback about the initial program. We downloaded and analyzed these electronic data to discern the nature of teacher and student experiences with the assessments. We analyzed and reported the 12 most relevant items to this study as the proportion (%) of respondents reporting favorable (either 1 or 2) responses. The Grade 4 to 5 teachers consisted of 11 males and 11 females, with 17 specialists (degrees in and certification in teaching PE) and five generalists (degrees and certification in another domain). The Grade 7 to 9 teachers consisted of 13 males and 18 females, with 26 specialists and five generalists.

Student profile. Prior to beginning PFL component assessments, students provided (once each year) general demographic data (e.g., gender, grade level) and responses to these two items: "How many times a week do you have PE class?" (0–5) and "How long are you physically active each day?" (1 = *less than 20 minutes*; 2 = *20 to 40 minutes*; 3 = *41 to 60 minutes*; 4 = *more than 60 minutes*). Because the second item closely reflects active participation, we reported it within that component in this study. For the administration of these and the other self-report items to Grade 3 students and to individuals with disabilities, teachers were encouraged to read and explain (as necessary) the items to students carefully. Because this demographic survey was optional, with only a portion of the students completing it, we provide their descriptive results only.

Living skills. The living skills component consisted of three short self-report surveys—to which participants responded on a 4-point scale (1 = *never*, 2 = *sometimes*, 3 = *most of the time*, 4 = *all of the time*)—that assessed feelings (seven items), thinking (seven items), and interacting skills (seven items) with positive links to healthy, active living. Table 2 provides a list of items by scale. The items representing the feeling scale assessed living skills related to students' motivation (participation, importance, enjoyment, self-efficacy, anxiety, body image, and autonomy) for physical activity.

Table 2
Living Skill Scale Items, Factor Loadings, and Variance Explained

	% of variance and factor loadings			
	Grades 3 to 6		Grades 7 to 9	
Scale, items, and item constructs	**2013–2014**	**2014–2015**	**2013–2014**	**2014–2015**
Feeling Scale, Items, and Item Constructs	45.63	53.85	54.53	52.29
1. I feel satisfied with the size and shape of my body. (Body Image) (4)	.53	.66	.74	.59
2. I feel that there are many physical activities that I can choose to do. (Autonomy) (8)	.74	.75	.56	.73
3. It is important to me to be physically active at least 60 minutes every day. (Importance Value) (9)	.69	.75	.76	.76
4. I feel relaxed (free of stress and fear) when performing physical activities. (Anxiety) (15)	.66	.71	.81	.68
5. I am sure that I can perform well in a wide variety of physical activities. (Self-Efficacy) (18)	.71	.78	.73	.80
6. I enjoy being moving and playing regularly. (Enjoyment Value) (21)	.71	.75	.80	.75

Table 2 (cont.)

Scale, items, and item constructs	% of variance and factor loadings			
	Grades 3 to 6		Grades 7 to 9	
	2013–2014	2014–2015	2013–2014	2014–2015
Thinking Scale, Items, and Item Constructs	42.07	48.08	47.21	45.30
1. I know and can use movement words (such as hop, spin, run, and leap) to explain how I move. (Conceptual Knowledge) (2)	.56	.64	.54	.76
2. Before I make a hard decision about something, I think carefully about all my choices. (Critical Thinking) (7)	.62	.61	.70	.57
3. I take time to think about my past mistakes in PE and try to learn from them so I can continue to improve in PE. (Critical Thinking) (11)	.70	.73	.73	.65
4. I can figure out how hard my body is exercising through simple ways like taking my own pulse or knowing how much I'm sweating. (Monitoring Knowledge) (12)	.70	.73	.69	.62
5. I am able to set goals for myself and meet them. (Goal Setting) (14)	.72	.75	.75	.73
6. I work hard to learn and enjoy physical activities even if I don't like them that much. (Effort Regulation) (16)	.63	.73	.76	.71
7. I understand how moving, playing regularly, and eating good food can make me healthier. (Understand) (17)	.60	.66	.62	.65

Table 2 (cont.)

Scale, items, and item constructs	% of variance and factor loadings			
	Grades 3 to 6		Grades 7 to 9	
	2013–2014	2014–2015	2013–2014	2014–2015
Interacting Scale, Items, and Item Constructs	46.17	46.48	47.84	45.85
1. When in a group, I am able to provide the time, effort, and support necessary to solve difficult problems. (Problem-Solving) (3)	.67	.67	.67	.74
2. When I'm safe but get really upset about something, I think it over or talk to someone about it rather than yell at or hit someone. (Self-Control) (5)	.55	.61	.66	.66
3. I am able to play with any other student in my class even if they have different beliefs than mine. (Respect) (6)	.64	.64	.66	.61
4. I can get along well with others in order to create a physical activity that is fun to play. (Cooperation) (10)	.70	.75	.62	.68
5. I care for other people by sharing with, playing fair with, including, and encouraging them. (Care/Empathy) (13)	.74	.74	.75	.66
6. I can use the right words to let someone know their actions are hurtful to me or others. (Assertiveness) (19)	.75	.63	.72	.68
7. I know where to find the information I need to help solve problems even if it means asking others for help. (Resource Management – Help Seeking) (20)	.70	.74	.75	.71

Note. Principal components exploratory factor analysis loading onto one factor for assessment Time 1. Items listed are for Grades 3 to 6.

Because one of these items ("I do at least 60 minutes of physical activity a day in which I am 'huffing and puffing' and feeling warmer") assessed participation in moderate to vigorous physical activity (MVPA), we moved it to the active participation component for this study. The items categorized under the thinking scale assessed students' relevant understanding, conceptual knowledge, monitoring knowledge, critical thinking, goal-setting, and effort-regulation behaviors for physical activity. Items in the relating scale reflected students' problem-solving, self-control, help-seeking, empathy, respect, assertiveness, and cooperation skills in physical activity settings.

Active participation. The active participation component consisted of 22 self-report items among three scales (diversity, interests, and intentions). The first 15 survey items investigated participation and interest in five movement domains, namely, team sport activities (e.g., hockey, soccer, volleyball, flag football), individual sport activities (e.g., golf, swimming, track and field, karate, cycling), dance activities (e.g., freestyle, hip-hop, ballet, jazz), gymnastics activities (e.g., trampoline, tumbling, rhythmic gymnastics), and fitness activities (e.g., jogging, yoga, body weight exercises, circuits, training). These three questions were asked for each domain: (1) "How often do you do activities at school (e.g., PE class, sports teams, clubs, recess, etc.)?" (2) "How often do you do activities at home or in the community (e.g., with friends, with family, sports teams, clubs, etc.)?" (3) "How interested are you in doing more activities?" The diversity scale consisted of the first two questions for each of these five movement areas (10 items) along with six items assessing students' level of participation in physical activities on/in land, snow/ice, water, air, and indoor and outdoor settings "at least once per week for more than four weeks in a row over the past year?" Response choices for this scale were 1 (*never*), 2 (*sometimes*), 3 (*most of the time*), or 4 (*all of the time*). The interests scale consisted of the third question (above) for each of the movement domains (totaling five items) and participants reported these on a 4-point scale: 1 (*not at all interested*), 2 (*a little interested*), 3 (*interested*), or 4 (*very interested*). Finally, we assessed intentions using the following item adapted from Prochaska and Velicer's (1997) stages of change model: "What statement best describes your intentions for physical activity?" Response choices for this item were on a 5-point escalating Likert scale.

Fitness skills. The fitness skills component consisted of three items. The Plank Challenge measured core muscle strength, the Lateral Bound assessed dynamic stability (balance), and the Four-Station Circuit provided evidence of aerobic (cardiorespiratory) endurance. For each item, teachers simultaneously assessed as many students as they deemed feasible. Teachers assigned each student a score using a 4-point scale (1 = *emerging*, 2 = *developing*, 3 = *acquired*, 4 = *accomplished*) based on detailed descriptions of each in a rubric provided to teachers.

Plank challenge. This item assesses the students' ability to use their core muscles (i.e., abdominals, lower back, hip) to maintain a static position. Students start in a "tabletop" position (i.e., elbows under the shoulders, knees under the hips) with their hands and elbows on the mat and their knees on the floor. The students' hands should form a triangular base but not be clasped together. Teachers prompt the students to start in a tabletop position and then to extend one leg and then the other leg to assume the elbow bridge position. Students should maintain this static position for as long as possible (up to 60 s), and they perform two trials with a 60-s break in between trials. Teachers monitor for flaws such as sagging in the lower back, any major movement to secure original position, raising the buttocks, major torso twists, bridging of the back, and/or dropping to the ground. A rubric for teachers describes that a score of 1 (*emerging*) should be assigned to students who use more than one rest interval or adjustment to maintain the proper position for the allotted time, a 2 (*developing*) to students who use one rest interval or adjustment, a 3 (*acquired*) to students who maintain the proper position for the allotted time of 60 s, and a 4 (*accomplished*) to students who do this for two consecutive trials.

Lateral bound. This item assesses the students' ability to consistently maintain a state of equilibrium in a static position after motion. Students attempt to balance on one leg, bound laterally across the required distance (sideways) landing on their opposite leg, and without pausing, bound back to their original leg and hold the position for at least 5 s so teachers can scan the group. Students are given one practice trial on each leg before testing, to allow them to determine their preferred leg. Students perform three trials and the teacher records the level that they achieve most consistently. Teachers use the scoring rubric to determine the level the student

achieves in each trial. The rubric informs teachers to watch for indications that students are unable to "stick" their landings, such as wobbling/wavering of the body, the other foot touching the ground, or the landing foot moving for students to gain a balanced position. For example, to earn a score of 4 (*accomplished*), students achieve a soft, balanced landing with correct triple flexion technique (landing leg is flexed at ankle, knees, hips; elbows close to body; head is up) without major adjustments occurring such as upper body wobbles, wavers, or twists; landing foot adjusting by swiveling or hopping; moving continuously to gain balance; or having the opposite foot touch the ground. On the opposite extreme, teachers assign a score of 1 (*emerging*) if participants are unable to bound across the required distance, fall down, adjust their body position more than once (full hop, major lower and upper body movement, both feet touch ground, other foot touches ground), or demonstrate continuous movement to achieve a balanced landing.

Four-station circuit. This item assesses students' ability to exercise at a vigorous intensity continuously for 9 min without walking, pausing, slowing down, or stopping. Each circuit is set up in an area the size of a badminton court. Students perform four stations (agility ladder, ball jumps, figure eights, and scissors) for 30 s each continuously (without a break) for as long as possible up to 9 min. For the agility ladder, students move in a hopscotch pattern (two-foot hop in first square, two feet straddling the ladder, two-foot hop into the second square, two feet straddling the ladder, etc.) up and then back through the ladder. For ball jumps, students start in a squat position to pick up the ball and jump, raising the ball over their heads (similar to a burpee jump). When the students land, they squat down to touch the floor with the ball and then jump up again. Students are reminded to land softly on both feet with both knees flexed to absorb the force. At the figure eights station, students stand between the two cones facing one side/alley of the badminton court. Students will know if they are doing this correctly if their hips and shoulders are always facing the same side/alley of the badminton court. Students move their feet forward and then backwards to take them in a figure eight pattern around the cones. Finally, at the scissors station, students face forward with one foot on each side of the badminton end line and switch their feet back and forth continuously.

Students should be using a contra-lateral (opposite) arm/leg movement with their feet contacting the ground at the same time and their body weight over the line. Throughout the assessment, the teacher scans the group by circulating in between the station lines and records indicators of fatigue on a sheet of paper every 30 s. Indicators of fatigue include walking between station changes, slowing down at a station, pausing or stopping, or losing their technique or stumbling. For example, students receiving a score of 1 (*emerging*) tend to use more than three rest intervals to engage in MVPA for less than 7 min, whereas those earning a 4 (*accomplished*) do so continuously for 9 min with no rest intervals.

Movement skills. As in the fitness skills component, the movement skills component consisted of three items that the teacher rated on a 4-point scale (1 = *emerging*, 2 = *developing*, 3 = *acquired*, 4 = *accomplished*) based on detailed descriptions of each in a teacher rubric. The Run-Stop-Return item measured locomotion, the Throw and Catch With a Bounce item assessed object control, and the Advanced Kick provided evidence of object manipulation.

Run-stop-return. This item assesses the students' ability to run, stop, and change direction in a controlled manner. Two cones are placed 7 m apart on a flat, clean surface free of obstacles or debris and at least 3 m from any wall before the start line and after the end line. A straight horizontal line is placed (with tape) at the start and finish, and another vertical line connects the two horizontal lines and cones. From a standing position behind the start line, the students (upon a prompt) run as fast as they can to the other line/cone. The students then stop, turn around, and run back. When the students stop, they should do so with control (i.e., without taking a lot of extra steps). Two trials are given with time provided between trials to enable the students to change something about their running technique based on what they have learned. The teacher observes for signs of clumsiness (e.g., tripping, stumbling), signs of lack of balance (e.g., flailing arms, sliding, falling down), and overall smoothness of the movement. For example, a score of 1 (*emerging*) reflects a student who stumbles or trips on the start or while running, is unable to stop in a controlled manner (e.g., overruns the line or takes extra steps when changing directions), and does not demonstrate a mature running pattern. Conversely, those earning a 4 (*accomplished*) sprint to the

line with a mature running pattern throughout, stop with control, and change directions smoothly with arms and legs.

Throwing and catching with a bounce. This item assesses the students' ability to throw a ball with accurate direction, velocity, and trajectory and to catch the ball that they threw. This task also challenges students to use problem-solving and cognitive skills to monitor how they should throw a ball to have it bounce back so they can catch it. A taped line is placed on the floor parallel to the wall that is 1 m away from the wall and 2 m long. A horizontal line is also placed on the wall that is 1 m away from the floor and 2 m long. The students stand anywhere they want behind the line but should not cross the line when throwing or catching the ball. The students throw the ball so that it bounces on the ground between the line and the wall and then hits the wall above the line on the wall. The students then try to catch the ball without stepping over the line. The activity is not timed and each student should perform the activity three times with the first trial being for practice. Teachers observe for (a) whether the ball bounces on the floor between the line and the wall (designated floor bounce), (b) whether the ball bounces above the line on the wall (designated wall bounce), (c) whether the student catches the ball, and (d) whether the student crosses the line on the floor when throwing or catching. For example, teachers give a score of 1 (*emerging*) if a student steps over the line to bounce the ball on the floor, hits the wall with the thrown ball but not above the line, and does not catch the ball. For students to earn a top score of 4 (*accomplished*), the ball bounces on the floor between the line and the wall and also bounces above the line on the wall. In addition, the students catch the ball with a controlled catch and adjust their body according to the trajectory of the ball.

Kicks. This item assesses the students' object manipulation skills using a task that involves the upper and lower body and a ball. This assessment involves two subtasks (a stationary kick and a punt). A taped line is placed on the floor running parallel to the wall, 4 m away from the wall, and 2 m long. A different horizontal line is placed on the wall that is 1 m away from the floor and 2 m long. For the stationary kick, the ball is placed on a cone or something else safe to hold it in place, 4 m from the wall. The students should kick the stationary ball above the line on the wall. If the students kick the ball over the

line on the first trial, they move on to the second task (punt). If not, they try the stationary kick a second time and do not move on to the second task (punt). For the punt, the students stand behind the line 4 m from the wall, hold the ball in their hands, and step forward and release the ball to punt it without letting the ball hit the ground first. The students are aiming to punt the ball above the line on the wall. If the students do not manage to punt the ball over the line on the first trial, the assessment is complete. If the students are able to punt the ball over the line on the first trial, they get a second attempt.

Compared to a participant earning a 4 (*accomplished*) evident in solid technique and outcome in both tasks, those earning a 1 (*emerging*) do not successfully perform the first task (stationary kick) either procedurally (e.g., stumble or use improper steps or foot contact) or in outcome (i.e., ball does not go over the line on any trial). Those receiving a 2 (*developing*) perform the stationary kick successfully with minor flaws but are not successful technically or in outcome on the second task (punt). A score of 3 (*acquired*) reflects that students are successful technically and in the outcome of the stationary kick yet inconsistent in their technical and outcome performance on the punt task.

Modifications for Grades 7 to 9

The PFL component assessments described for Grades 3 to 6 were identical for Grades 7 to 9 except for minor differences in the living skills and fitness assessments and more extensive alterations in the movement skills. The duration of the Four-Station Circuit (Cardiovascular Endurance) assessment was 12 min (9 min for Grades 3 to 6) and the width between lines for the Lateral Bound (Dynamic Stability) was 85 cm (75 cm for Grades 3 to 6). In terms of living skills, 17 of the 20 items were either identical or had only a simple word(s) replaced to increase their complexity and relevance (e.g., "hard decision" to "difficult decision") for Grades 7 to 9. The meaning of the remaining five items was maintained despite altering the wording for complexity and relevance (e.g., "I feel that there are many physical activities that I can choose to do" vs. "I feel that I have many choices and options about the physical activities I can participate in").

Each of the movement skill assessments was significantly different for Grades 7 to 9 than for Grades 3 to 6. The Throw and Catch

(Object Control) assessment was the least modified by requiring the participant to stand and throw from a distance of 3.5 m from the wall (1 m for Grades 3 to 6) and to throw off the wall first (off the floor first for Grades 3 to 6) so that the ball bounces into an area 2 to 2.75 m from the wall before the participant attempts to catch it (while still standing behind the line 3.5 m from the wall). The teacher rubric was altered accordingly with *accomplished* as the "ball bounces in target area." The student catches the ball with one hand in a controlled manner without major body movement. The student does not take steps to catch the ball. The Kicks (Object Manipulation) assessment involving both a place and punt kick in Grades 3 to 6 was changed to the Punt and Catch for Grades 7 to 9 to assess students' object control and manipulation skills using a task that involves the upper (catch) and lower (punt) body and a ball. In the task, participants demonstrate the coordinated hand and foot motion to control kick (punt) a ball and predict motion of a ball to pursue, intercept, and catch it. The student drops a soccer ball held in both hands and punts (kicks it before it hits the ground) from behind a 3-m line and catches the ball after it bounces off the wall while never crossing the 3-m line. Finally, Run-Stop-Return (Locomotion) movement skill assessment was changed to Run-Side Shuffle-Back Pedal (Locomotion) to reflect more diverse means of locomotion in Grades 7 to 9.

The setup involves placing four cones—one in each corner of a rectangular (7 m × 3 m) flat surface and at least 3 m from any object (e.g., wall). From a position slightly left of the first cone and on a teacher prompt, the participants run as fast as they can up to and slightly past the pylon 7 m distant, then perform a rightward side shuffle to the adjacent pylon 3 m to the right, and then back pedal toward the final pylon 7 m to the rear and stop. Using the teacher rubric for guidance, the instructor observes for strong acceleration and deceleration, control, and fluid conversion between running, shuffling, and back pedaling.

Results

Content Evidence

Development. In this section, we add to and elaborate on the content validity evidence provided in the Method section. The

development of PFL involved an extensive consultation process. In January 2011, PHE Canada invited 15 to 20 PE experts from across Canada to meet to discuss the potential of a physical literacy assessment tool to be used in schools across Canada and, if recommended, what the guiding principles might be. Based on the outcomes of that meeting, in the summer of 2011 PHE Canada organized a gathering of six Canadian professor-researchers with expertise in PE curricula and pedagogy, physical literacy, and assessment, along with several school-based PE teacher-educator leaders, to develop PFL based on the guiding principles and its definition of physical literacy. This led to the formulation of the names and basic composition of the four components of PFL (Active Participation; Movement, Fitness, and Living Skills). Cognition was not a component, because it was viewed as being integrated and evident in the four components. Five members of this group then wrote and developed each of the component assessments and also met occasionally to review, plan, and provide guidance on PFL. This writing team represented a balance of expertise in each PFL component from across Canada. Each writer also had practical teaching-coaching experiences with children and adolescents in physical activity settings, four were certified specialists in physical and health education, and three were professor-researchers with terminal degrees in a relevant field (e.g., PE). Several iterations of each assessment were made based on feedback from the writers.

The results of the pilot test with Grade 4 and 5 students revealed that PFL data by measure were normally distributed (e.g., no ceiling or floor effect evident in skew or kurtosis) and had satisfactory concurrent validity (e.g., theoretically expected correlations among constructs and measures). For example, Pearson bivariate correlations among and between the three movement skill assessments and three fitness skill assessments were positive and statistically significant ($p < .001$) ranging from .28 to .45. Additionally, these skills correlated positively and significantly ($p < .01$) to students' self-reported participation levels in fitness activities at school ($r = .11–.20$). A comparison sample of 20 students from a fifth grade class was used to test the interrater and test–retest reliability and provide validity evidence of concurrent relations with PFL fitness and movement skill measures.

The intraclass correlation coefficient among three simultaneous raters using PFL rubrics for fitness and movement skills revealed good scoring agreement (.65 to .82). The analysis of test–retest reliability using the data from the three raters 1 week later revealed strong consistency (r = .72 to .89). Based on the feedback from this pilot test, modifications were made (most notably to the content of the living skills items) and expanded to include Grade 3 students with the Grades 4 to 6 assessment for full implementation in the fall of 2013–2014, along with a new PFL for Grades 7 to 9. Following the pilot of PFL with Grades 4 to 5 in 2012–2013 and revisions stemming from it, PFL was administered to Grades 3 to 6 and Grades 7 to 9 in 2013–2014 and 2014–2015.

Components. It was beyond the scope of this study to postulate much about the nature and content of physical literacy beyond highlighting that the four components of PFL and their assessments reflect the holistic nature of physical literacy as defined by PHE Canada (2015), SHAPE America (Roetert & MacDonald, 2015), and others (e.g., Whitehead, 2001). It is important, however, to demonstrate validity evidence that the content of PFL assessments adequately reflect the four component constructs (active participation and living, fitness, and movement skills) of PFL and that these represent the fundamental aims and outcomes of PE. Evidence for this is solid because of the extensive consultation process involving numerous experts in the relevant fields (e.g., assessment, PE, physical literacy) during the development of PFL, which was designed to align with vital psychomotor, health-related physical activity participation and fitness, affective, and cognitive learning outcomes embedded in PE curricula across Canada (Kilborn et al., 2016) and elsewhere (Lacy & Hastad, 2003; SHAPE America, 2014). Support for the four PFL components and their assessments is also evident in positive empirical associations between fundamental motor skills, physical activity and fitness, and advantageous psychosocial qualities, particularly during adolescence (e.g., Lubans, Morgan, Cliff, Barnett, & Okely, 2010; Stodden et al., 2008; Weiss, 2011).

By component, active participation in PFL was designed to reflect evidence that physically literate individuals intentionally attain regular and suitable physical activity at necessary levels of moderate to vigorous intensity. For example, current guidelines recommend

that children and adolescents engage in at least 60 min of MVPA every day (Canadian Society for Exercise Physiology, 2011; U.S. Department of Health and Human Services, 2009). To meet such guidelines, active participants tend to engage in and show interest in diverse physical activities (e.g., team, individual, dance, educational gymnastics, and fitness) in different settings (e.g., indoors and outdoors, at home, at school, and in the community) and environments (e.g., on ice, snow, and land, and in the water and air; Haga, 2008; Roetert & MacDonald, 2015).

The three assessments for living skills (feeling, thinking, and interacting) reflect noteworthy psychosocial characteristics with positive empirical links to being physically active. These include being motivated for and having advantageous feelings (e.g., body image, self-efficacy, interest, enjoyment); demonstrating relevant knowledge, understanding, critical thinking, and goal-setting behaviors; and having relational qualities such as problem-solving, personal and resource management, and communication skills (e.g., for reviews, see Chen, 2001; Weiss, 2011). Because of the quantity of psychosocial correlates of physical activity and the need for a concise survey to minimize time and increase rate of participation, we used one item to assess each of the psychosocial constructs (see Table 2). We modified these items slightly from previously used measures of the constructs to accommodate the Grade 3 to 6 and 7 to 9 developmental levels. For example, the stem of the item assessing self-efficacy ("I am sure that I can perform well in a wide variety of physical activities") has been used in many other diverse learning settings (Pintrich, Smith, Garcia, & McKeachie, 1991). Further, despite some compromises to psychometric properties, single-item scales are suitable for use in assessments such as living skills in PFL even for empirical research (Jordan & Turner, 2008).

The two teacher-administered PFL components (fitness and movement skills) represent two pillars of PE curricula (Newell, 2011) and were designed in PFL to provide affirming yet challenging physical assessments that might help youth to be aware of and interested in their levels of fitness and movement proficiency for optimal engagement in a lifetime of play, activity, and exercise. Gallahue, Ozmun, and Goodway (2012) reported that fundamental movement skills (categorized into stability, locomotor, and manipulative/object

control) provide the physical foundation or building blocks for developing more complex physical proficiencies and skills used in games, activities, sports, and leisure pursuits. The three PFL movement skill assessments of running (locomotor), kicking (object manipulation), and throwing and catching (object manipulation) reflect these core movement competencies that are commonly assessed in youth (e.g., Vandaele, Cools, de Decker, & Martelaer, 2011) and associated with healthy, active outcomes (Lubans et al., 2010; Stodden et al., 2008).

Research evidence suggests that children can develop these fundamental movement skills by age 6.5, yet a surprisingly low proportion of children do so (Gallahue et al., 2012). In regard to PFL fitness assessments (muscle strength, balance/dynamic stability, and cardiorespiratory endurance), they represent core aspects of fitness in PE (e.g., Kilborn et al., 2016; Lacy & Hastad, 2003; MacKenzie, 2003; SHAPE America, 2014). These fitness skills have also been linked to healthy, active outcomes and reflect more basic movement proficiencies. For example, each PFL fitness assessment reflects a person's cognitive understanding and postural control (i.e., balance and the ability to maintain a desired body orientation), which are prominent rate limiters of early motor development (Clark, 2007).

Response Process Evidence

We assessed whether the PFL component scores reflect the construct, particularly through the way teachers administer the assessments and participants engage with the items. The evidence consisted of teacher feedback on the exit survey following the Grade 4 to 5 pilot study and the first year of study with Grades 7 to 9. On the basis of their experience with PFL, the majority of teachers (95% for Grades 4 to 5; 83% for Grades 7 to 9) felt that students had a better understanding of physical literacy and that students' results would help improve their physical literacy (91% for Grades 4 to 5; 97% for Grades 7 to 9). Over 93% of the teachers of students in Grades 7 to 9 reported that students were easily able to understand and complete the assessments for each of the components, with none reporting strong disagreement. This sentiment was similar among Grade 4 to 5 teachers for the fitness and movement skills (> 85%), yet substantially lower yet satisfactory for living skills (71%) and for active participation (66%). The most common reason for these lower ratings

for the self-report assessments in Grades 4 to 5 was the difficulty many of the students had in comprehending the items. For example, one teacher responded, "Some students needed help with the wording of some statements as to what it meant."

Another important aspect of the response process evidence, particularly for criterion-referenced practitioner-based assessments for educational settings such as PFL, is its administrative usability for teachers, because "a careless approach to test administration can result in invalid scores" (Lacy & Hastad, 2003, p. 190). Results of the exit survey revealed that the teachers were generally positive about aspects of usability such as accessibility, relevance, and ease of using PFL teachers' guide with instructions (100% for Grades 4 to 5; 90% for Grades 7 to 9) and scoring rubrics (81% for Grades 4 to 5; 97% for Grades 7 to 9) for informing their teaching (100% for Grades 4 to 5; 97% for Grades 7 to 9). Teachers also reported that it was easy for them (100% for Grades 4 to 5; 84% for Grades 7 to 9) and their students (94% for Grades 4 to 5; 81% for Grades 7 to 9) to navigate the PFL website. The most prevalent administrative concern was the amount of class time spent completing PFL with 43% reporting that it took an unreasonable amount of time. According to most of the teachers, completing PFL required between 2.5 and 6 classes, which is reasonable, particularly given that psychometric qualities tend to be compromised on shorter tests (Gosling, Rentfrow, & Swann, 2003), particularly tests that assess constructs such as physical literacy, active participation, and fitness, living, and movement skills (Lacy & Hastad, 2003). Several teachers also expressed the need for more video aids to help with the movement and fitness skills and to assist them with particular challenges in administering some of the assessments (e.g., four-station cardiovascular circuit) simultaneously to small groups of students. In response, PHE Canada has added a variety of video tutorials and demonstrations.

Internal Structure Evidence

The descriptive statistics (i.e., means, standard deviations, skew, and kurtosis) revealed no abnormalities in any of the items or scales, particularly for large sample sizes (Tabachnick & Fidell, 2006), except for a consistently (by year and developmental level) negative skew and kurtosis (> 1.00) for the *intentions to be physically active* item

in active participation. Consequently, this item was not included in subsequent analyses. Another source of validation evidence is the degree that scores from the assessment items align (interrelate) with the construct being assessed relative to the intended interpretation of the assessments (Standards; AERA, APA, & NCME, 2014). We used principal component exploratory factor analyses loading onto a single factor and suppressing factor loadings < .30 to explore the factor structure for each living skill scale (feeling, thinking, and interacting) by year and level (Grades 3 to 6, Grades 7 to 9).

Because of the large sample sizes for the first assessment of each year relative to those lower than recommended for factor analysis (< 250) in the second assessment of each year (Tabachnick & Fidell, 2006), we used only the first assessment by year and level. Table 2 shows the results. Each item in each scale had strong factor loadings (.53–.81), which explained a satisfactory proportion of the variance (42.07–54.53%). Especially for scales with fewer than 10 items (Loewenthal, 1996), the internal consistency reliability coefficients (see Table 3) for each of the three scales by year and level were satisfactory (living skills, .75 to .86; active participation, .61 to .87; movement skills, .64 to .75; fitness skills, .62 to .77).

Noting the likelihood of maturation and other effects (e.g., motivational and seasonal) between repeated assessment times in the fall and spring seasons within each academic year (2013–2014, 2014–2015) and developmental level (Grades 3 to 6, Grades 7 to 9), we computed and reported test–retest reliability coefficients for the assessment scales to potentially signal the stability of measurement scores. We aligned the criterion with Loewenthal's (1996) standards for scales with fewer than 10 items, setting it at >.43 for an interitem test–retest correlation coefficient and > .60 for an internal consistency reliability coefficient.

Table 3
Descriptive Statistics and Internal Consistency for PFL Components by Assessment, Level, Year, and Time

Components, scales, and (# of items)	Grades 3–6								Grades 7–9							
	2013–2014				2014–2015				2013–2014				2014–2015			
	T1		T2		T1		T2		T1		T2		T1		T2	
	M (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α
Movement Skills (3)	2.66 (.77)	.67	3.22 (.72)	.78	2.64 (.74)	.64	2.89 (.75)	.73	2.51 (.71)	.64	3.00 (.64)	.65	2.66 (.74)	.68	3.02 (.77)	.75
1. Locomotion	2.82 (.82)	_	3.23 (.73)	_	2.60 (.80)	_	2.95 (.75)	_	2.40 (.88)	_	2.80 (.76)	_	2.66 (.83)	_	3.09 (.78)	_
2. Object Control	2.70 (1.07)	_	3.33 (.86)	_	2.81 (.97)	_	3.03 (.97)	_	2.68 (.94)	_	3.20 (.77)	_	2.78 (.95)	_	3.10 (.97)	_
3. Object Manipulation	2.42 (1.05)	_	3.11 (.99)	_	2.54 (1.09)	_	2.71 (1.02)	_	2.44 (.99)	_	3.01 (.95)	_	2.58 (1.03)	_	2.90 (1.03)	_
Fitness Skills (3)	2.39 (.86)	.70	3.00 (.81)	.77	2.34 (.81)	.63	2.69 (.75)	.62	2.22 (.83)	.73	2.75 (.80)	.73	2.38 (.83)	.67	2.78 (.78)	.68
1. Balance	2.60 (.90)	_	3.27 (.75)	_	2.47 (.89)	_	2.96 (.76)	_	2.47 (.94)	_	2.97 (.82)	_	2.58 (.87)	_	3.13 (.78)	_
2. Muscle Endurance	2.44 (1.16)	_	2.91 (.07)	_	2.26 (1.01)	_	2.65 (1.09)	_	2.35 (1.17)	_	2.87 (1.00)	_	2.34 (1.10)	_	2.77 (1.00)	_
3. Cardiorespiratory Endurance	2.19 (1.12)	_	2.81 (1.09)	_	2.19 (1.05)	_	2.55 (1.04)	_	1.84 (1.03)	_	2.42 (1.09)	_	2.15 (1.09)	_	2.51 (1.09)	_
Living Skills (20)	3.14 (.49)	.90	3.16 (.54)	.93	3.04 (.54)	.93	3.15 (.52)	.92	3.01 (.52)	.92	3.09 (.53)	.93	3.08 (.52)	.92	3.08 (.52)	.92
1. Feeling (6)	3.26 (.55)	.75	3.29 (.62)	.86	3.03 (.62)	.83	3.17 (.61)	.83	3.01 (.65)	.83	3.09 (.64)	.84	3.12 (.62)	.81	3.12 (.63)	.83
2. Thinking (7)	3.08 (.54)	.77	3.08 (.59)	.84	3.12 (.56)	.82	3.22 (.55)	.83	2.97 (.56)	.81	3.08 (.56)	.84	3.05 (.56)	.80	3.05 (.54)	.80
3. Interacting (7)	3.10 (.56)	.86	3.16 (.59)	.84	2.98 (.58)	.83	3.06 (.57)	.81	3.04 (.54)	.84	3.11 (.55)	.85	3.07 (.55)	.83	3.08 (.55)	.81
Active Participation (22)	2.52 (.51)	.87	2.59 (.53)	.89	2.40 (.52)	.88	2.51 (.53)	.89	2.35 (.52)	.89	2.41 (.52)	.89	2.42 (.51)	.88	2.55 (.47)	.86
1. Diverse PA (16)	2.41 (.52)	.84	2.49 (.52)	.86	2.31 (.52)	.85	2.41 (.55)	.87	2.25 (.51)	.85	2.31 (.52)	.85	2.33 (.51)	.85	2.44 (.49)	.84
2. Interest (5)	2.59 (.69)	.68	2.57 (.72)	.73	2.42 (.68)	.69	2.54 (.64)	.61	2.41 (.68)	.70	2.51 (.68)	.71	2.44 (.65)	.66	2.58 (.62)	.61
MVPA/Day (1)	2.82 (.86)	_	3.02 (.88)	_	2.89 (.89)	_	3.05 (.83)	_	2.83 (.92)	_	2.94 (.89)	_	2.95 (.90)	_	3.03 (.88)	_

Note. All on a 4-point scale. MVPA = moderate to vigorous physical activity.

No items in the data for any year or level were below an interitem test–retest correlation coefficient of .24 or an internal consistency reliability coefficient of .40. In the Grade 7 to 9 samples, each of the items, scales, and components met the stability criterion with the exception of balance (.24 and .38) in 2013–2014 and several living skill items (five in 2013–2014, two in 2014–2015). In other words, 84–94% of PFL assessment items (41–46 of 49) met the criterion. This was somewhat lower in the two assessment times among the Grade 3 to 6 samples with 67% (33 of 39) meeting the criterion over the 2 years. Quantities of items by year that exceeded the criterion among the Grade 3 to 6 students were active participation (two in 2014–2015, 10 in 2013–2014), living skills (12 in 2014–2015, six in 2013–2014), fitness skills (balance in 2014–2015), and movement skills (object control in 2014–2015). Of course, item differences by assessment period might also be due to the different seasons when each assessment was administered (the fall and winter), hence their fluctuating responses to items that ask them to report, for example, their MVPA and frequency of participating in activities in water or on the ice/snow.

Pearson bivariate correlations between each scale within each component (Tables 4 and 5) were generally reflective of conceptual and theoretical expectations based on previous research. Scale correlations within each PFL component construct (e.g., feeling, thinking, and interacting scales within the living skills component) for the first assessment time in all four data sets had positive and significant associations ($p < .01$) and showed no evidence of problematic multicollinearity (> .80). Moreover, the correlations between each of the three scales within each component in all four data sets (2 years for two grade levels) were statistically significant ($p < .01$) with values ranging from .32 to .47 (movement skills), .63 to .79 (living skills), .31 to .52 (fitness skills), and .19 to .70 (active participation). These relations are further reflected in the significant ($p < .01$) between-component and MVPA correlations, as presented in Table 5. These results, along with the satisfactory internal consistency reliability coefficients for each scale, support the internal structure of each component and its scales.

Table 4
Component Scale Correlations by Year (2013–2014, 2014–2015) and Grade Level (3–6, 7–9)

Scale (component)	1 2013–14 (2014–15)	2 2013–14 (2014–15)	3 2013–14 (2014–15)	4 2013–14 (2014–15)	5 2013–14 (2014–15)	6 2013–14 (2014–15)	7 2013–14 (2014–15)	8 2013–14 (2014–15)	9 2013–14 (2014–15)	10 2013–14 (2014–15)	11 2013–14 (2014–15)	12 2013–14 (2014–15)
1. MS1 (Locomotion)	–	.41* (.33*)	.32* (.35*)	.41* (.38*)	.36* (.28*)	.48* (.32*)	.21* (.25*)	.17* (.19*)	.10* (.14*)	.18* (.22*)	.02 (.18*)	.17* (.21*)
2. MS2 (Throw-Catch)	.34* (.37*)	–	.47* (.45*)	.29* (.24*)	.30* (.26*)	.35* (.11)	.17* (.10)	.10* (.08)	.04 (.04)	.13* (.08)	.01 (-.01)	.14* (.09)
3. MS3 (Kick-Punt)	.31* (.43*)	.46* (.45*)	–	.28* (.22*)	.30* (.24*)	.36* (.07)	.16* (.05)	.10* (.00)	.03 (.00)	.15* (.11*)	-.07 (.02)	.21* (.02)
4. FS1 (Balance)	.35* (.41*)	.21* (.31*)	.16* (.20*)	–	.38* (.37*)	.41* (.34*)	.14* (.12*)	.13* (.13*)	.10* (.09)	.17* (.16*)	.12* (.15*)	.10* (.12*)
5. FS2 (Core Muscle Endurance)	.39* (.40*)	.23* (.29*)	.21* (.21*)	.42* (.46*)	–	.52* (.37*)	.14* (.14*)	.13* (.11*)	.03 (.13*)	.23* (.18*)	.07 (.12*)	.14* (.13*)
6. FS3 (Cardiorespiratory Endurance)	.47* (.41*)	.20* (.18*)	.14* (.18*)	.41* (.31*)	.51* (.44*)	–	.20* (.17*)	.12* (.15*)	.05 (.13*)	.18* (.19*)	.01 (.13*)	.14* (.16*)
7. LS1 (Feeling)	.30* (.27*)	.22* (.09)	.25* (.19*)	.11* (.23*)	.31* (.18*)	.29* (.22*)	–	.71* (.79*)	.64* (.74*)	.48* (.48*)	.33* (.37*)	.47* (.68*)
8. LS2 (Thinking)	.19* (.17*)	.16* (.04)	.18* (.09)	.08* (.18*)	.23* (.11*)	.21* (.09)	.76* (.77*)	–	.74* (.75*)	.50* (.39*)	.35* (.32*)	.46* (.59*)
9. LS3 (Interacting)	.12* (.10)	.10* (-.03)	.09* (.01)	.06 (.10)	.17* (.02)	.14* (.06)	.63* (.67*)	.78* (.79*)	–	.43* (.37*)	.30* (.26*)	.37* (.50*)
10. AP1 (PA Diversity)	.28* (.30*)	.18* (.15*)	.18* (.18*)	.11* (.24*)	.23* (.23*)	.25* (.33*)	.55* (.44*)	.52* (.42*)	.39* (.35*)	–	.64* (.73*)	.44* (.41*)
11. AP2 (Diverse Interest)	.18* (.17*)	.11* (.04)	.08* (.06)	.10* (.19*)	.17* (.19*)	.18* (.19*)	.44* (.32*)	.46* (.36*)	.40* (.31*)	.70* (.70*)	–	.19* (.28*)
12. MVPA	.24* (.18*)	.16* (.16*)	.17* (.12*)	.06 (.20*)	.23* (.11*)	.25* (.12*)	.65* (.61*)	.60* (.62*)	.45* (.48*)	.53* (.36*)	.39* (.30*)	–

Note. Grades 3–6 in top diagonal and Grade 7–9 in bottom diagonal. Time 1 only. MS = movement skill; FS = fitness skill; LS = living skill; MV = moderate to vigorous physical activity.

*$p < .01$.

Table 5
Component and Physical Activity Correlations by Year (2013–2014, 2014–2015) and Grade Level (3–6, 7–9)

Components	1 2013–14 (2014–15)	2 2013–14 (2014–15)	3 2013–14 (2014–15)	4 2013–14 (2014–15)	5 2013–14 (2014–15)
1. Movement Skills (3)	–	.52* (.39*)	.17* (.13*)	.18* (.16*)	.23* (.13*)
2. Fitness Skills (3)	.42* (.44*)	–	.15* (.19*)	.22* (.23*)	.15*(.17*)
3. Living Skills (20)	.26* (.15*)	.26* (.16*)	–	.53* (.46*)	.48* (.64*)
4. Active Participation (22)	.29* (.26*)	.27* (.33*)	.57* (.46*)	–	.42* (.41*)
5. MVPA (1)	.25* (.19*)	.23* (.17*)	.53* (.62*)	.55* (.38*)	–

Note. Grades 3 to 6 in top diagonal and Grades 7 to 9 in bottom diagonal. Time 1 only. MVPA = moderate to vigorous physical activity.

*$p < .01$.

Relations to Other Variables

A fourth source of validation evidence is the degree that subscale scores and/or their correlations relate to similar or associated variables concurrently or predictively measured by other instruments (AERA, APA, & NCME, 2014). Because this was a descriptive validation study of data from the initial uses of PFL, we did not statistically compare scores on PFL assessments to those on other previously validated instruments; however, we report validation evidence based on how each PFL scale and component relate concurrently. We also provide validation evidence as to the temporal (predictive) consistency in relations between the first and second year of data (2013–2014, 2014–2015) for each scale by developmental level.

Scale and component relations. As illustrated in Table 4, the majority of relations in both years of data for both developmental levels were significantly positive ($p < .01$). Across all data sets (level and year), the strongest scale relations existed between each of the movement and fitness skills, and between each of the living skills (feeling, thinking, and interacting) and active participation scales (PA diversity, diverse interest, and MVPA). Associations between the scales of the first two components (movement and fitness skills) and the latter two components (movement and fitness skills) were mixed. For example, locomotion was positively associated with all of the other scales except for diverse interest in 2013–2014 for the

Grade 3 to 6 sample and interacting in 2014–2015 for the Grade 7–9 sample. Relations between each of the scales and MVPA were also generally positive and significant ($p < .01$), although links between MVPA and active participation and living skills were notably higher than with fitness and movement skills. This pattern of scale relations was reinforced in the between-component and MVPA correlation, as presented in Table 5.

Temporal consistency of scale and component relations. None of the bivariate correlations for either year within Grades 7 to 9 were statistically negative, and 65 of the 66 paired correlations (e.g., relations between two scales in 2013–2014 compared to 2014–2015) differed by less than .15. This was slightly lower across the 2 years of Grade 3 to 6 data with 60 (90.9%) of the correlations differing by less than .15, and 47 of the 66 pairs (71%) being positive and statistically significant ($p < .01$). This temporal consistency was also evident in the relations between the components and MVPA, as depicted in Table 5, with each being statistically positive ($p < .01$) and 18 of the 20 paired correlations (90%) differing by less than .15. This suggests substantial temporal (predictive) consistency of the correlations across each scale and component over the 2 years of testing for both developmental levels.

Discussion

This paper served as a descriptive validity analysis for the use of PFL as a feasible formative criterion-referenced practitioner-based assessment of physical literacy. In line with the Standards (AERA, APA, & NCME, 2014), four broad sources of validation evidence were provided from the 2 years of PFL data in Grades 3 to 6 and Grades 7 to 9. First, score validity evidence based on the content of PFL included the development of PFL by a number of domain experts, the nature and format of the components and scales relative to existing literature and evidence, and the administrative procedures to guide teachers to implement it. Second, internal structure evidence attested to how well the instrument-specific items and their meanings aligned with each component construct and scale through indicators of normality, internal consistency reliability, factor structure, and test–retest reliability. Third, response process evidence included teachers' feedback, which generally demonstrated that the scores appear to reflect the construct, particularly through the way

that participants engaged with and understood the items. Finally, we explored relations with other variables within PFL and found substantial positive and significant within- and across-component relations. Substantial temporal (predictive) consistency in relations between scales and components over the 2 years of testing for both developmental levels was also evident. The results of this study of initial validation evidence support the delivery of PFL for its intended use, but we note several cautions and recommendations.

Critical in any validation analysis is a full consideration of the contextual purpose of the assessment. In essence, PFL is a series of assessments that represent cornerstone construct components of a broader physical literacy construct that is increasingly aligned with PE standards and outcomes (Mandigo et al., 2009; Ontario Ministry of Education, 2015; Roetert & MacDonald, 2015). In other words, the content of PFL aligns with PE outcomes and criterion standards that can be used to promote positive aspects of physical literacy such as learning, positive attitudes, and goal setting. Teachers are informed of these important PFL principles; particularly they should know and remind students that the assessments are for learning and improvement of each student's physical literacy and that students are comparing themselves to a criterion standard of physical literacy for their age group and not normatively to other students. The need to balance necessary psychometric properties with administrative usability qualities such as accessibility, relevance, financial and duration costs, clarity of instructions and scoring rubrics, and quality of resources such as a teachers' guide is also important. With the exception of a few thematic concerns such as duration and challenges in understanding some self-report items in the lower grades, teachers were generally satisfied with the administrative usability of PFL for themselves and their students.

The results from this study lend further credibility of the holistic nature of physical literacy. Similar to literacy being the ability of students not only to read and write but also to read and interact with their culture and environment (Whitehead, 2001), physical literacy is more than simply the physical skills needed to move. It involves the constant interaction between a person's physical abilities (e.g., movement and fitness skills); their motivation and desire to be physically active across a number of activities in multiple environments (e.g.,

active participation); and their confidence and competence to apply their skills, knowledge, and attitudes for themselves and with others on a daily basis (e.g., living skills). Most of the measures representing the various components of physical literacy in PFL correlated positively and significantly; this was particularly so between movement and fitness skills and between active participation and living skills. Although correlations between the physical measures and the affective and cognitive measures were somewhat mixed (particularly for younger participants), many of these relations were significantly positive. Collectively, these associations support the holistic nature of physical literacy consisting of psychomotor, health-related physical activity participation and fitness, and affective and cognitive learning outcomes embedded throughout PE curricula across Canada (Kilborn et al., 2016) and elsewhere (Lacy & Hastad, 2003; SHAPE America, 2014).

Because of the relative infancy of PFL, additional studies are needed to provide more robust validation evidence relative to its intended use. Intentional validation studies that provide much less duration between repeated measures (for test–retest reliability), test concurrent and predictive relations with relevant constructs using other measures, and compile ratings from a group of random experts in PE and physical literacy on the content relevance of PFL are needed.

A more rigorous analysis of each self-report item (e.g., think-aloud protocols), particularly relative to the lower grades, would also be useful. It appears that a significant portion of students may struggle with understanding the questionnaire items, which influences the response accuracy and hence the validity of those items and their respective scales and components (living skills and active participation). Future studies should also provide validation evidence pertaining to the fifth broad source of validation evidence (AERA, APA, & NCME, 2014), which was not included in this study, that is, linking the results of PFL to future intended or unintended consequences. Less stability in the test–retest reliability coefficients of several items in active participation and living skills for the Grade 3 to 6 sample may signal that younger students have difficulty accurately calibrating and reporting on these. Future PFL studies can also assess test–retest reliability of tests in closer tem-

poral proximity while investigating grade-level specificity responses by item and scale, particularly in Grades 3 to 6. Finally, although beyond the scope of this initial validation study, it would be useful to perform confirmatory factor analysis of the three living skills surveys, repeated measures analysis of variance procedures to determine change by year and grade, and structural equation modeling to better illustrate the nature and direction of relationships among the PFL constructs.

To add to the validation evidence for the intended use of PFL, it is important that the participant samples adequately represent the constituent members. These initial PFL data lack balanced representation across Canada (particularly the northern regions with a larger Inuit and Aboriginal population), and adequate accommodations have yet to be provided to individuals with special needs. There is also a lack of information on each participant's age, gender, province, ethnicity, grade, disability, and residence (urban or rural). A more robust scale for physical activity and MVPA also appears necessary, particularly for use as a criterion from which to compare each component and its scales.

PHE Canada has noted some of these needs (T. Zakaria, personal communication, November 11, 2015) and has recently increased the amount of demographic information that participants must report, has pilot tested an enhanced PFL program with optional adaptations embedded to accommodate individuals with particular needs (i.e., disabilities), and has moved participants in Grade 3 from the 3 to 6 program to a new program for kindergarten to Grade 3 students set to begin in September 2017. Among the adaptations is a simplified version of the self-report (living skills and active participation) items. Based on this study, it also appears necessary to revise the items on MVPA, intentions, and frequency of participating in individual sports and fitness activities at school or at home and in the community for Grades 3 to 6 because they may have difficulty relating to the wording of these items and accurately responding. There is also potential confirmability bias in this study because the lead author was involved in the design of several of the PFL assessments. Construct underrepresentation is also possible in PFL because each item can only reflect, not equate, with its intended scale or component or with physical literacy itself.

In conclusion, with an increased focus on the importance of physical literacy within PE settings, there is more of a need to provide students with formative feedback while they progress along their physical learning journey. Developed specifically for education settings and for educators, this study has provided initial validity evidence to support the continued use of PFL in Canadian schools. Like physical literacy, the concept of validity is a lifelong journey. Although no tool is free of measurement error and there are always areas for improvement as highlighted in this study, there is sufficient evidence to suggest that PFL is on the right path to helping inform Canadian children and adolescents of their current levels of physical literacy.

Armed with this information, parents, teachers, and students themselves will be in a better position to support children and adolescents. Further, children and adolescents will be better equipped to make informed decisions and become more physically literate individuals.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Canadian Institutes of Health, Natural Science, and Engineering Research Council of Canada & Social Sciences and Humanities Research Council of Canada. (2014). *Tri-council policy statement: Ethical conduct for research involving humans.* Ottawa, Canada: Government of Canada.

Canadian Society for Exercise Physiology. (2011). *Canadian physical activity guidelines for youth - 12–17 years.* Retrieved from http://www.csep.ca/CMFiles/Guidelines/CSEP-InfoSheets-youth-ENG.pdf

Chen, A. (2001). A theoretical conceptualization for motivation research in physical education: An integrated perspective. *Quest, 53,* 35–58. https://doi.org/10.1080/00336297.2001.10491729

Clark, J. E. (2007). On the problem of motor skill development. *Journal of Physical Education, Recreation, and Dance, 78*(5), 39–44. https://doi.org/10.1080/07303084.2007.10598023

Gallahue, D. L., Ozmun, J. C., & Goodway, J. D. (2012). *Understanding motor development* (7th ed.). New York, NY: McGraw-Hill.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37,* 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Haga, M. (2008). The relationship between physical fitness and motor competence in children. *Child Care, Health, and Development, 34,* 329–334. https://doi.org/10.1111/j.1365-2214.2008.00814.x

Jordan, J. S., & Turner, B. A. (2008). The feasibility of single-item measures for organizational justice. *Measurement in Physical Education and Exercise Science, 12,* 237–257. https://doi.org/10.1080/10913670802349790

Kilborn, M., Lorusso, J., & Francis, N. (2016). An analysis of Canadian physical education curricula. *European Physical Education Review, 22*(1), 23–46. https://doi.org/10.1177/1356336X15586909

Lacy, A. C., & Hastad, D. N. (2003). *Measurement and evaluation in physical education and exercise science* (4th ed.). Toronto, Canada: Pearson Education.

Loewenthal, K. M. (1996). *An introduction to psychological tests and scales.* London, England: UCL Press.

Lubans, D. R., Morgan, P. J., Cliff, D. P., Barnett, L. M., & Okely, A. D. (2010). Fundamental movement skills in children and adolescents: Review of associated health benefits. *Sports Medicine, 40,* 1019–1035. https://doi.org/10.2165/11536850-000000000-00000

Mandigo, J., Francis, N., Lodewyk, K., & Lopez, R. (2009). Physical literacy for educators. *Physical and Health Education Journal, 75*(3), 27–30.

Newell, K. M. (2011). Physical education of and through fitness and skill. *Quest, 63,* 46–54. https://doi.org/10.1080/00336297.2011.10483662

Ontario Ministry of Education. (2015). *The Ontario curriculum grades 1 to 8: Health and physical education.* Toronto, Canada: Queens Printer.

Physical and Health Education Canada. (n.d.). What is physical literacy? Retrieved August 11, 2017, from http://www.phecanada.ca/programs/physical-literacy/what-physical-literacy

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)* (Technical Report No. 91-B-004). Ann Arbor: University of Michigan, School of Education.

Prochaska, J. O., & Velicer, W. F. (1997). The trans-theoretical model of health behavior change. *American Journal of Health Promotion, 12*(1), 38–48. https://doi.org/10.4278/0890-1171-12.1.38

Roetert, E. P., & MacDonald, L. C. (2015). Review: Unpacking the physical literacy concept for K–12 physical education: What should we expect the learner to master? *Journal of Sport and Health Science, 4,* 108–112. https://doi.org/10.1016/j.jshs.2015.03.002

SHAPE America. (2014). *National standards and grade-level outcomes for K–12 physical education.* Champaign, IL: Human Kinetics.

Sport Canada. (2012). *Canadian sport policy 2.0.* Ottawa, ON: Government of Canada.

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest, 60,* 290–306. https://doi.org/10.1080/00336297.2008.10483582

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics.* Boston, MA: Allyn & Bacon.

United Nations Educational, Scientific, and Cultural Organization. (2003). *Literacy, a UNESCO perspective.* Paris, France: Author.

U.S. Department of Health and Human Services. (2009). *2008 physical activity guidelines for Americans.* Retrieved from http://www.health.gov/paguidelines/

Vandaele, B., Cools, W., de Decker, S., & de Martelaer, K. (2011). Mastery of fundamental movement skills among 6-year-old Flemish pre-school children. *European Physical Education Review, 17*(1), 3–17. https://doi.org/10.1177/1356336X11402268

Weiss, M. R. (2011). Teach the children well: A holistic approach to developing psychosocial and behavioral competencies through physical education. *Quest, 63,* 55–65. https://doi.org/10.1080/00336297.2011.10483663

Whitehead, M. (2001). The concept of physical literacy. *British Journal of Teaching Physical Education, 32*(1), 6–8. https://doi.org/10.1080/1740898010060205