

A Construct Validation Study of PLAYfun

JOHN CAIRNEY^{1,2}, SCOTT VELDHIJZEN², JEFFREY D. GRAHAM², CHRISTINE RODRIGUEZ¹,
CHLOE BEDARD^{2,3}, EMILY BREMER^{2,4}, and DEAN KRIELLAARS⁵

¹Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON, CANADA; ²Infant and Child Health (INCH) Research Lab, Department of Family Medicine, McMaster University, Hamilton, ON, CANADA; ³Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, CANADA; ⁴Department of Kinesiology, McMaster University, Hamilton, ON, CANADA; and ⁵College of Rehabilitation Sciences, University of Manitoba, Winnipeg, MB, CANADA

ABSTRACT

CAIRNEY, J., S. VELDHIJZEN, J. D. GRAHAM, C. RODRIGUEZ, C. BEDARD, E. BREMER, and D. KRIELLAARS. A Construct Validation Study of PLAYfun. *Med. Sci. Sports Exerc.*, Vol. 50, No. 4, pp. 855–862, 2018. **Introduction:** Physical Literacy Assessment for Youth (PLAYfun) is a measure of motor competence, comprehension, and confidence which is part of a suite of scales used to assess physical literacy in children and youth; however, its measurement properties have not been reported in the published literature. The purpose of this study is to examine the factor structure of PLAYfun, in addition to variations in PLAYfun subscale results by age and sex. **Method:** In this study, we use a sample of children and youth 7 to 14 yr of age ($n = 215$) to test a proposed factor structure for the motor competence component of PLAYfun and to examine age and sex differences in subscale and total scores. The initial ($n = 128$) and secondary ($n = 98$) samples were drawn from a stratified (by geographic region), random sample of 27 after-school programs from a larger pool of 400 programs across the province of Ontario. Seven research assistants were initially trained on the administration of PLAYfun and rated a small pilot sample of 10 children. These trained assessors then assessed the full sample. **Results:** Interrater agreement was very good (intraclass correlation, 0.87). The hypothesized five-factor structure of the scale was found to have an acceptable fit to the data (root mean square error of approximation, 0.055; 90% confidence interval, 0.03–0.075; comparative fit index, 0.95; Tucker–Lewis Index, 0.94). In general, PLAYfun scores increased with age as developmentally expected. There were few sex differences across skills, but girls did not perform as well as boys on upper and lower body object control skills. **Conclusions:** The factor structure and patterns of results by age and sex support PLAYfun as a measure of motor competence. Continued evaluation of the tool and other subscales of PLAY is required. **Key Words:** PHYSICAL LITERACY, MEASUREMENT, CHILDREN, YOUTH, MOTOR COMPETENCE

Physical inactivity is a major public health concern (1,2). One factor that is posited to increase participation in physical activity, particularly in children and youth, is an individual's level of movement competence—with those individuals who are more competent engaging in higher levels of physical activity than their less competent peers (3,4). However, our understanding of the associations between movement competence and physical activity is limited by our ability to assess movement competence. Most movement skill assessments are designed for children and youth and do not discriminate well between levels of competency; rather, they tend to dichotomize whether a skill is possessed or not, or whether developmental components of a skill have been achieved (e.g., Test of Gross Motor Development-2 (TGMD-2) [5]). Given the

potential for movement competence to influence an individual's participation in lifelong physical activity, it is important that we are able to measure it using a tool that can capture the variability in movement competencies that exist across the life course.

Physical literacy is a multidimensional construct composed of both physical and psychological attributes theorized to be foundational to participation in physical activity (6,7). Although many different definitions exist, most include movement competencies (8), positive affective states (fun, enjoyment), motivation (e.g., perceived competence, confidence, motivation), and knowledge of the importance of movement to health and well-being (9–11). Given the inclusiveness and comprehensiveness of the construct, physical literacy has been positioned as an important part of intervention and policy in health, recreation, physical education, and sport (12). Furthermore, viewing movement competence within the comprehensive framework of physical literacy may allow us to better understand the variability in actual competence over time.

Enthusiasm for physical literacy, though, has outpaced empirical validation of the construct. At present, there are only a few assessment tools specifically identified as measures of physical literacy (13–15). Of these, only the Canadian Assessment of Physical Literacy (CAPL) has been the focus of some initial peer-reviewed psychometric research (14). In this article, we examine the construct validity of another

Address for correspondence: John Cairney, Ph.D., Faculty of Kinesiology and Physical Education, University of Toronto, 55 Harbour St, Toronto, ON, Canada M5S 2W6; E-mail: john.cairney@utoronto.ca.

Submitted for publication August 2017.

Accepted for publication October 2017.

0195-9131/18/5004-0855/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2017 by the American College of Sports Medicine

DOI: 10.1249/MSS.0000000000001494

instrument, the Physical Literacy Assessment for Youth fun (PLAYfun) tool, which measures fundamental, land-based movement competencies that are physical education curriculum linked (e.g. Refs. [16,17]), and hypothesized to be essential to physical literacy in children and youth 7 yr and older. PLAYfun is part of a larger suite of measures that together form the PLAY tools (13). Together, these measures provide a comprehensive assessment of physical literacy, measuring competency in movement skills (PLAYfun), and affective states and behavior (PLAYself), consistent with the definition of the construct provided by Whitehead (10) and others (9,18). The PLAY tools also include assessment measures for coaches (PLAYcoach), physical education teachers (PLAYpe), parents (PLAYparent), and recently movement creativity (PLAYcreativity). Given the centrality of motor competence to the construct of physical literacy, we limit our focus in this article to assessment of PLAYfun.

PLAYfun differs from existing measures of physical literacy (e.g., CAPL) and other assessments of fundamental movement skills both in terms of content (e.g., specific skills measured) and in terms of scoring. With regard to the latter, the system for rating each skill uses a modified visual analog scale (VAS), which requires raters to assess children and youth along a continuum. This is quite different from other measures that use, for example, product-based outcomes (e.g., time to complete a task) such as the Movement-Assessment Battery for Children (19) or the Bruiniks–Oseretsky Test of Motor Proficiency (20), and/or process evaluation measures that rate the quality of a child's movement, usually using a 2-point ordinal scaling system (e.g., TGMD-2 [5]). A challenge with the latter scaling approach is that variability is potentially reduced when only a small number of response options are provided. At least in theory, the scaling system used in PLAYfun should produce greater variability and should not demonstrate ceiling effects, which should lead to higher reliability of the scale, greater discriminate validity, and increase the strength of correlations with other measures (21).

In this study, we examine the factor structure of PLAYfun and examine variations in PLAYfun subscale results by age and sex. Although PLAYfun (and the other scales that comprise PLAY) have been endorsed by the Sport4Life organization for the assessment of physical literacy and are being used by practitioners in many different settings in Canada and the United States (e.g., recreational centers, after-school programs, sport organizations), at present, there is no published evidence of its construct validity.

METHODS

Sample. We selected a stratified, random sample of 27 after-school programs from a larger pool of 400 programs across the province of Ontario.

We used programs' postal codes to identify their location within Ontario. We then divided them into regions: the Greater Toronto Area (GTA), and the Northern, Eastern, and Southwestern areas of the province. For logistical reasons, we randomly

selected 20 sites from the GTA, which were within 2-h travel of our laboratory. We then randomly chose another seven sites from the remainder of the province, including at least one from each of the Northern, Eastern, and Southwestern regions.

Programs had to have a reported minimum of 20 participating children to be eligible. We excluded smaller programs to limit the number of sites necessary to obtain an adequate sample size.

Because of the wide range of ages served by some programs, however, it proved necessary to recruit further sites. We therefore drew eight further sites, of which one was in Southwestern Ontario and seven in the GTA. This ensured that a pool of 300–400 children 8 to 14 yr of age was available for the study. However, the number of participants who completed assessments from this initial sampling ($n = 128$) was lower than expected and, more importantly, lower than sample size estimates for confirmatory factor analysis ($N = 200–250$). As such, we randomly sampled another nine sites from the GTA, which yielded a sample of 98 participants. Our final sample included 226 consented participants. Of these, assessments were not done for 8 (3.7%) and were incomplete for 3 others (1.4%). The remaining 215 children comprise the final sample.

Sample size recommendations for confirmatory factor analysis with maximum likelihood vary. Simulation work suggests that a total sample between 200 and 250 is usually adequate (22,23), but some research also suggests that fit problems and fit index performance depend on complex interactions and the characteristics of data and models. Our sample of 215 satisfies some sample size recommendations, but not all. However, our model was fit with no difficulties, and all parameters were estimated successfully.

Site recruitment. Sites randomly selected from the pool of after-school programs were contacted and invited to participate in the study. Of the 27 sites selected, 3 were excluded because ethics approval from their school board was required and could not be obtained within the timeframe of the study: one site was no longer part of the program and had been replaced by another site, which was invited to participate in the present study, and finally, we were unable to enroll any parents or children at 2 sites. The organization coordinating one of the sites offered a different site in the same city to replace it. This alternate site was also invited to participate in the study. In the end, 24 sites across the province of Ontario participated.

Study design. This was a cross-sectional study in which children were assessed using the PLAYfun tool. For the initial sample, parents and children were recruited between February and April 2016, and assessments were conducted between March and May 2016. For the second sample, parents and children were recruited between September and November 2016, and assessments were conducted in December 2016. PLAYfun was administered by graduate students and research assistants, all of whom had an undergraduate or master's degree (e.g., kinesiology, health sciences). According to the training manual, it is to be used by trained professionals (i.e., coaches, exercise professionals or individuals trained in the analysis of movement) only. Before testing, all assessors completed more than 10 h of training, including an orientation

session led by the designer of the measure, Dr. Dean Kriellaars. We also examined interrater agreement before commencement of the study by having seven assessors assess a small sample of subjects ($n = 10$) in a fully crossed design (all participants were assessed by each tester, in-person during a single session; scoring was completed independently). The study was approved by an institutional research ethics board. All parents provided informed, written consent and all children provided informed, written assent.

Measure. PLAYfun is one tool from the PLAY collection of tools to measure physical literacy in individuals 7 yr and older. PLAYfun comprises 18 different movement tasks within five domains that assess different aspects of a child's movement skills. The five domains are as follows: 1) running, 2) locomotor, 3) object control—upper body, 4) object control—lower body, and 5) balance, stability, and body control. The tasks included within each domain are provided in Table 1.

Before the start of the assessment, participants are given a general set of instructions to explain that they will be asked to perform a number of movements and that they should try to perform to the best of their ability. A concise instruction is then provided to the participants before each individual skill is performed (e.g., “I want you to run a square around the pylons. I want you to run a square as best you can. Ready? Run now”); there is no skill modeling provided to the participants. Participants perform a single trial for each of the 18 tasks, and the complete assessment takes approximately 15 min to complete (13).

Children are assessed using a VAS that is 100 mm in length and divided into four categories: initial (0–25 mm),

emerging (25–50 mm), competent (50–75 mm), and proficient (75–100 mm). The initial and emerging categories are indicative of an individual who is still developing a given skill, whereas, the competent and proficient categories indicate that the skill has been acquired. Once the assessor determines within which category the child falls, the assessor places an “X” within the 25-mm box of the category to be more specific when denoting the child's skill for the particular task. For example, an individual who has just acquired a given skill may be placed in the lower end of the competent category, whereas someone who displays a greater level of competence would be placed higher in that category. Individuals in the proficient category would typically have a high level of skill-specific training through sport or other activities (e.g., dance, gymnastics, etc.). The PLAYfun training manual provides detailed exemplars for each category for each of the 18 tasks. The scale is a holistic rubric and not criterion based. Children are not scored relative to other children of the same age; instead, the 100-mm scale represents all people regardless of age. The top score (i.e., 100) is defined as “the very best anyone could be at the skill regardless of age.” To derive the score for each task, a ruler is used to measure from the beginning of the scale (i.e., developing) to the center of the “X.” Thus, each task is given a score of 0–100. The total score is the sum of all 18 task scores, and domain scores are the sum of scores of tasks included within the domain.

The birth date and sex of the child were recorded for each participant. Age was calculated by subtracting the date of testing from the date of birth for each child.

TABLE 1. Item and subscale descriptive statistics for PLAYfun ($n = 215$).

	Mean	SD	Min	Max	Skewness	Kurtosis
Running						
Run a square (T1)	48.8	10.1	8.0	72.0	-0.9	4.7
Run there and back (T2)	53.3	8.3	26.5	73.0	-0.4	3.7
Run, jump then land on 2 feet (T3)	48.2	12.5	1.5	75.0	-0.9	4.6
Locomotor						
Crossovers (T4)	28.5	22.2	1.0	77.0	0.7	2.1
Skip (T5)	46.4	15.4	1.5	73.0	-1.2	3.6
Gallop (T6)	39.2	17.2	2.5	73.0	-0.4	1.9
Hop (T7)	51.7	11.1	5.5	77.0	-1.0	4.8
Jump (T8)	42.4	16.6	3.0	76.0	-0.6	2.5
Object control—upper body						
Overhand throw (T9)	51.2	13.4	8.5	76.5	-0.7	3.6
Strike with a stick (T10)	50.1	12.3	13.0	80.0	-0.3	3.5
One-handed catch (T11)	49.5	20.8	1.0	87.0	-0.4	2.0
Hand dribble stationary and moving Forward (T12)	51.4	13.4	1.5	80.0	-1.0	4.7
Object control—lower body						
Kick ball (T13)	50.7	16.0	5.5	80.5	-0.7	3.1
Foot dribble moving forward (T14)	44.7	16.8	1.5	77.0	-0.9	3.5
Balance, stability, and body control						
Balance walk (heel-to-toe) forward (T15)	46.8	19.0	0.5	80.5	-1.0	3.1
Balance walk (toe-to-heel) backward (T16)	43.1	18.0	1.0	80.0	-0.9	3.0
Drop to the ground and get back up (T17)	44.9	14.3	1.0	86.0	-0.3	3.3
Lift and lower (T18)	56.7	17.4	1.0	80.5	-1.0	3.6
Hypothesized subscales						
Running (T1–T3)	50.1	8.3	22.5	71.2	-0.3	3.9
Locomotor (T4–T8)	41.7	11.5	10.3	73.8	-0.1	3.2
Object control (upper) (T9–T12)	50.5	11.2	16.9	79.0	-0.2	2.9
Object control (lower) (T13–T14)	47.7	14.4	4.0	77.0	-0.7	3.3
Balance (T15–18)	47.9	13.0	7.4	81.8	-0.8	3.5
Total (T1–T18)	47.1	9.3	14.8	72.9	-0.3	3.7

T, task.

Analysis. We measured interrater agreement using intraclass correlation (ICC) on our pilot sample of 10 children. With the full sample ($n = 215$), we first produced descriptive statistics of the PLAYfun tasks and examined their central tendency, variance, skewness, and kurtosis. To evaluate the published factor structure of the instrument, we used confirmatory factor analysis. Because PLAYfun is only one component of a comprehensive measure of physical literacy, we chose a correlated traits model to assess fit to the data. Our model is based on the PLAYfun manual, which groups tasks explicitly into five domains: 1) running, 2) locomotor, 3) object control—upper body, 4) object control—lower body, and 5) balance, stability, and body control. We tested this hypothesized factor structure, allowing all factors to be correlated with the others, with tasks treated as manifest indicators. We fit models using maximum likelihood estimation.

Given concerns about the use of modification indices to improve model fit (22), we considered potential modifications only in cases where the theoretical case seemed exceptionally clear-cut. To evaluate model fit, we followed the guidelines of Hu and Bentler (22), who suggest interpreting model fit in terms of the comparative fit index (CFI), Tucker–Lewis Index (TLI), and root mean square error of approximation (RMSEA), with values greater than 0.9 as “acceptable” and greater than 0.95 as “good” for the TLI and CFI, and values less than 0.08 as “acceptable” and less than 0.05 as “good” for the RMSEA.

Finally, we examined age- and sex-dependent variation in PLAYfun domain and total scores. We used Pearson correlation for age and *t*-tests for sex differences. For subscales with significant sex differences, we calculated effect sizes (Cohen’s *d*). We used STATA version 14 for all analyses.

RESULTS

The ICC coefficient for the total score among seven assessors in our pilot sample of 10 children was very good (ICC, 0.87).

Our final sample included 112 (52%) boys and 103 (48%) girls. Age was not recorded for one child. For the remaining 214, the average age was 10.3 yr (SD, 1.7), with a minimum of 6.5 and a maximum of 14.1. The average number of children registered in the participating programs was 39.3 (SD, 9.0).

Table 1 reports descriptive statistics for each item in the PLAYfun measure. In general, the range of scores indicates that raters were using most of the scale to evaluate each child, but no child was rated higher than 87. This is in fact expectable because the end point range (approximately 75 and higher) is reserved for elite-level athletes. The median and mean values for most items are close to 50, which is the midpoint of the scale; scores for crossovers and gallop (tasks 4 and 6, respectively), however, were somewhat lower. The values for skewness indicate that most items have distributions that are reasonably symmetrical. Measures of kurtosis indicate that items 1, 2, 3, and 7 showed somewhat peaked distributions. In all cases, however, departures from normality were not dramatic. Table 2 is a correlation matrix for all items, subscales, and PLAYfun total scores. Item-total correlations varied from 0.47 to 0.83, a suitable range considering developmental acquisition and requirement for practice for achievement of motor proficiency. In general, correlations of less than 0.4 are described as weak, those from 0.4 to 0.59 as moderate, and those of 0.6 or higher as strong.

The fit of the initial model was fair (RMSEA, 0.065; 90% confidence interval, 0.052–0.077; CFI, 0.93; TLI, 0.91). Modification indices suggested several ways the model could be adjusted, but most would have represented *post hoc* adjustments without truly clear and obvious justifications. One change, however, was clearly reasonable: adding a path to allow error terms for tasks 15 and 16 to covary. These items load on the same factor and more importantly are categorically identical (body control and balance), with the direction of movement (forward and backward) being the

TABLE 2. Item and subscale correlations.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1. #1	1.00																								
2. #2	0.58	1.00																							
3. #3	0.39	0.48	1.00																						
4. #4	0.37	0.45	0.35	1.00																					
5. #5	0.29	0.29	0.28	0.35	1.00																				
6. #6	0.26	0.31	0.27	0.29	0.35	1.00																			
7. #7	0.44	0.44	0.53	0.39	0.36	0.38	1.00																		
8. #8	0.44	0.45	0.44	0.32	0.30	0.31	0.55	1.00																	
9. #9	0.35	0.46	0.41	0.30	0.23	0.22	0.36	0.44	1.00																
10. #10	0.39	0.48	0.35	0.28	0.23	0.24	0.34	0.28	0.48	1.00															
11. #11	0.26	0.39	0.34	0.33	0.14	0.08	0.31	0.29	0.32	0.38	1.00														
12. #12	0.34	0.41	0.51	0.27	0.16	0.21	0.42	0.36	0.44	0.40	0.44	1.00													
13. #13	0.23	0.42	0.25	0.20	0.21	0.16	0.31	0.25	0.38	0.37	0.42	0.41	1.00												
14. #14	0.31	0.50	0.39	0.37	0.15	0.26	0.34	0.35	0.43	0.42	0.39	0.53	0.54	1.00											
15. #15	0.42	0.35	0.30	0.29	0.29	0.24	0.46	0.44	0.38	0.27	0.25	0.41	0.32	0.38	1.00										
16. #16	0.38	0.35	0.33	0.29	0.29	0.26	0.48	0.47	0.38	0.27	0.25	0.45	0.29	0.38	0.90	1.00									
17. #17	0.31	0.43	0.37	0.33	0.24	0.19	0.30	0.29	0.25	0.33	0.46	0.39	0.31	0.34	0.22	0.22	1.00								
18. #18	0.17	0.24	0.22	0.16	0.29	0.17	0.33	0.29	0.21	0.16	0.23	0.27	0.28	0.32	0.42	0.44	0.26	1.00							
19. Run	0.79	0.81	0.82	0.47	0.35	0.34	0.59	0.55	0.50	0.49	0.41	0.53	0.36	0.49	0.44	0.44	0.45	0.26	1.00						
20. Loc	0.51	0.56	0.52	0.73	0.67	0.67	0.71	0.69	0.44	0.39	0.33	0.39	0.31	0.43	0.47	0.49	0.39	0.34	0.65	1.00					
21. OCU	0.44	0.58	0.53	0.40	0.25	0.23	0.47	0.45	0.71	0.71	0.79	0.75	0.53	0.58	0.43	0.44	0.49	0.29	0.63	0.51	1.00				
22. OCL	0.31	0.53	0.37	0.33	0.21	0.24	0.37	0.35	0.46	0.45	0.46	0.54	0.87	0.88	0.40	0.38	0.37	0.34	0.49	0.42	0.64	1.00			
23. Bal	0.43	0.45	0.40	0.35	0.37	0.28	0.52	0.50	0.41	0.34	0.38	0.50	0.39	0.47	0.87	0.88	0.52	0.71	0.52	0.56	0.54	0.49	1.00		
24. Total	0.60	0.70	0.63	0.60	0.50	0.47	0.69	0.66	0.63	0.58	0.59	0.66	0.58	0.68	0.69	0.69	0.56	0.51	0.79	0.83	0.82	0.72	0.81	1.00	

Bal, balance; Loc, locomotor; OCL, object control—lower; OCU, object control—upper; Run, running.

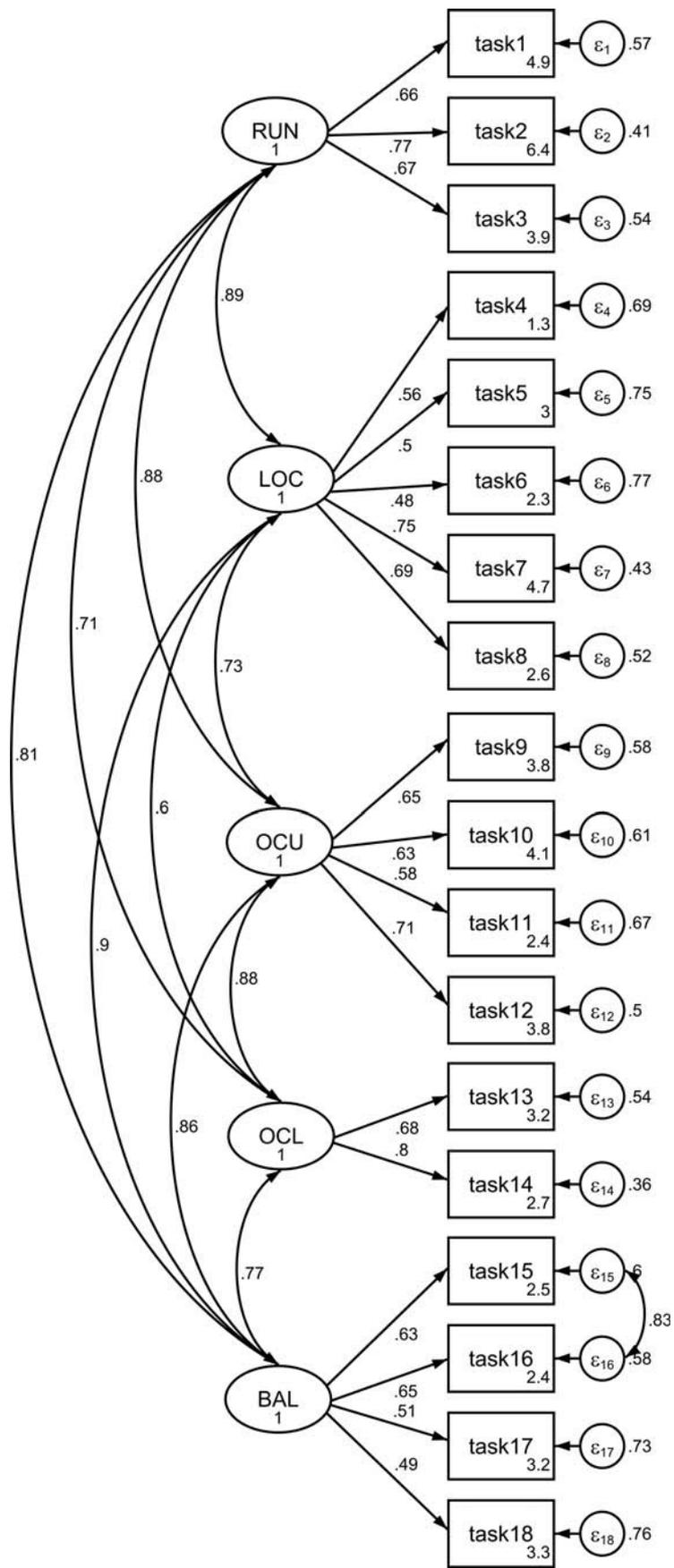


FIGURE 1—Factor structure of PLAYfun.

only distinction. As it seemed clear that a particularly close relationship could be expected between these items, we added this path and refit the model. The adjusted model is shown in Figure 1. This modification improved fit indices somewhat (RMSEA, 0.055; 90% confidence interval, 0.03–0.075; CFI, 0.95; TLI, 0.94).

Having established the five-factor model for PLAYfun, we examined descriptive statistics for each of the five subscales (see Table 1). The values for the mean and median suggest that most children are rated at the midpoint of the scales for each of the items under each domain. The values for skewness and kurtosis are close to 0 and 3, respectively, indicating that the distributions for each subscale are reasonably normally distributed.

Table 3 reports sex differences for both the subscales and the total PLAYfun scale. Boys scored marginally higher overall, but this difference was relatively small and was not significant. The only substantial and significant differences in subscales were the two “object control” subscales; on both of these, boys scored slightly higher than did girls, with Cohen’s *d* of 0.49 for “object control—upper” and of 0.36 for “object control—lower.” Other subscales also slightly favored boys, with the exception of “locomotor,” which was marginally and nonsignificantly higher among girls.

Table 4 shows the correlations between age and each of the subscales and total scale scores for the full sample, and separately for boys and girls. All correlations are positive and moderate in size, indicating that older children, on average, perform better on PLAYfun and on all subscales compared with younger ones. Correlations with age were slightly stronger for boys on running and locomotor domains, but were otherwise similar.

DISCUSSION

This is the first study to examine core psychometric properties of the PLAYfun tool. In particular, we examined two different aspects of construct validity: factor structure and variation in scores due to age and sex. The latter analysis was based on previous literature showing differences between sex and age on fundamental movement skills in children and youth in this general age range (24). Confirming the factor structure of a scale is an important aspect of validity because it is important to assess both the overall dimensionality of scales and whether the hypothesized structures of scales can be confirmed empirically (21). Given the increasing interest

TABLE 3. Sex differences in total score and subscales on PLAYfun.

	Male (<i>n</i> = 112), Mean (SD)	Female (<i>n</i> = 103), Mean (SD)	Total (<i>n</i> = 215), Mean (SD)	<i>t</i> (213)	<i>P</i>
Running	51.1 (9.4)	49.1 (6.6)	50.1 (8.2)	1.84	0.07
Locomotor	40.9 (12.5)	42.4 (10.2)	41.6 (11.4)	1.07	0.28
Object control (upper)	53.3 (11.9)	47.8 (9.7)	50.6 (11.2)	3.63	<0.01
Object control (lower)	50.1 (15.2)	44.9 (13.1)	47.6 (14.4)	2.49	0.01
Balance	46.6 (15.2)	49.2 (9.9)	47.9 (13)	1.6	0.11
Total	47.7 (10.7)	46.5 (7.3)	47.1 (9.2)	0.78	0.44

TABLE 4. Correlations with age on PLAYfun total score and subscales.

	Male (<i>n</i> = 112)		Female (<i>n</i> = 103)		Full Sample (<i>N</i> = 215)	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>
Running	0.51	<0.001	0.26	0.01	0.42	<0.001
Locomotor	0.52	<0.001	0.24	0.02	0.41	<0.001
Object control (upper)	0.57	<0.001	0.55	<0.001	0.53	<0.001
Object control (lower)	0.45	<0.001	0.51	<0.001	0.46	<0.001
Balance	0.54	<0.001	0.37	0.001	0.48	<0.001
Total	0.63	<0.001	0.50	<0.001	0.58	<0.001

r, Pearson’s *r*.

in physical literacy and the relative paucity of both measures and research on the measurement of priorities of existing tools, including PLAYfun, this study fills an important gap in the current literature in this field.

Overall, the hypothesized factor structure of the scale, which consisted of five domains (running, object control—upper body, object control—lower body, locomotor, and balance), was empirically supported using confirmatory factor analysis. These results confirm that PLAYfun is a multidimensional scale measuring multiple aspects of land-based movement competencies, which are key competencies that are linked to health and physical education curricular expectations for children (16,17).

PLAYfun was designed to test individuals 7 yr and older. In general, we expect performance on the movement competency tasks included in PLAYfun to improve with age, as movement skill in general has been found to improve with age in normative samples (25,26). A positive correlation between age and PLAYfun scores, therefore, would support one important aspect of construct validity. Our results confirm that the aggregate and subscale scores of PLAYfun, on average, do indeed improve with age.

Sex differences were generally small and, in most instances, nonsignificant, except in relation to object control skills, both upper and lower body: our results show that boys were rated higher on these tasks than girls. That boys perform better on skills involving hands, feet, and objects (e.g., balls, bats) is consistent with previous literature (24). One explanation for sex differences in object control skills is that difference is experiential: boys are more likely to play games and sports that involve object control compared with girls and therefore have greater opportunity to develop proficiency in these skills (27–29). In addition, there may be sex biases in delivery of recreation and physical education programming in provision of opportunity. Although the findings relating to the progression of movement skills (i.e., running and locomotor) with age being stronger for boys is troublesome, it also highlights an important area for future interventions targeting this sex difference. Nevertheless, it is important that PLAYfun is able to detect sex differences, because this information can be used both for evaluation and for the planning of programs targeting skill improvement.

As noted at the outset, the VAS ratings used in PLAYfun should lead to greater range in movement skill assessments, especially when compared against other measures of FMS that use limited, ordinal response rating options (e.g., TGMD). These criterion-based assessments do not allow for variability in how a task is performed outside of the traditional motor

sequences established as being “proficient” or “mature” for a given skill. This means that individuals do not receive credit for proficiency above an entry-level standard, nor do they receive credit for performing a skill with form that is only partially correct from a motor development perspective. Furthermore, older children tend to reach a ceiling in these other criterion-based assessments as they all perform in a proficient manner, albeit at varying levels of proficiency that is not discerned by an ordinal response. Ceiling effects present in motor skill assessments that are age referenced such as the TGMD or the Bruiniks–Oseretsky Test of Motor Proficiency have been widely identified (5,20,30–32). VAS ratings, such as the ones used in PLAYfun, may address the problem of ceiling effects by increasing range in response options. Indeed, our results show that raters in this study used almost the full range of the VAS scales when evaluating children and youth suitable for this age range (where we would not expect expert mastery meaning few scores greater than 85). At the same time, the end point of “very best anyone could be” will, by design, lead to end point aversion. Indeed, not one child or youth in this study received a score of 100 on any of the tasks. Future studies should examine children and young adults (14–25 yr of age) in sport-specific settings to examine the ability to reach the upper 25 of the scale.

It is important to note that PLAYfun is only part of the PLAY tools, which measures behavioral and affective states as well as movement skill. Although physical literacy is more than just movement skill, the latter is nevertheless a core element of the construct, and evaluation of the movement skill battery in PLAYfun is therefore a reasonable place to begin. Further work will need to examine the remaining subscales, including examination of the factor structure of the whole system, and should compare results with those obtained using CAPL. CAPL uses a circuit to assess “coordinated” movement skill sequences rather than individual movement skills (e.g., catching a ball with one hand), and a comparison of these approaches would be of considerable interest. Comparing these two measures is desirable because PLAYfun is both easier and quicker to administer than the circuit assessment in CAPL. It will also be important to compare PLAYfun to other movement

skill batteries (e.g., TGMD, for the range that TGMD does not exhibit ceiling effects), which would provide evidence of convergent validity. Given the very good characteristics of the PLAYfun scale, research comparing PLAYfun to physical activity, ideally using objective measures (e.g., accelerometers), will be important to conduct. Other measures of motor competency have shown low correlations with objectively measured PA (33), the characteristics of the PLAYfun scale may provide greater prediction without the limitation of other scales. Finally, it would be useful to assess whether PLAYfun is sensitive to change, which would require longitudinal and/or intervention designs.

Several limitations should be acknowledged. First, our sample size was not as large as some recommended minimums for conducting confirmatory factor analysis (34). It is important to replicate this analysis in a different and larger sample to increase confidence in the results. As noted previously, we were not able to compare PLAYfun with other measures of motor competence, nor did we have additional data on children (e.g., habitual physical activity) to assess other aspects of construct validity. Finally, although sites were randomly selected, participants were not.

This is the first published study to examine important aspects of construct validation in relation to PLAYfun. Given that this measure is endorsed by a national sport organization and is being used already for the purposes of assessing physical literacy in children and youth, these results make an important and overdue contribution to our understanding of the measure. That PLAYfun seems to be measuring what it was hypothesized to measure, at least in relation to factor structure, and its ability to differentiate children by age and sex (object control subscales) is encouraging.

This work was supported by funding from Sport for Life Society, Ontario Trillium Foundation, and the Government of Ontario.

The authors declare no potential conflict of interest, and that the results of the present study do not constitute endorsement by the American College of Sports Medicine. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. Kohl HW, Craig CL, Lambert EV, et al. The pandemic of physical inactivity: global action for public health. *Lancet*. 2012;380(9838):294–305.
2. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380(9838):219–29.
3. 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(12):1019–35.
4. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. *Sports Med*. 2015;45(9):1273–84.
5. Ulrich DA. *Test of Gross Motor Development*. 2nd ed. Austin (TX): Pro-Ed; 2000.
6. Higgs C, Balyi I, Cardinal C, Norris S, Bluechardt M. *Developing Physical Literacy: A Guide for Parents of Children Ages 0 to 12*. Vancouver (BC): Canadian Sports Centres; 2008.
7. Tremblay M, Lloyd M. Physical literacy measurement—the missing piece. *Phys Health Educ J*. 2010;76(1):26.
8. Dudley DA. A conceptual model of observed physical literacy. *Phys Educ*. 2015;72:236–60.
9. Edwards LC, Bryant AS, Keegan RJ, Morgan K, Jones AM. Definitions, foundations and associations of physical literacy: a systematic review. *Sports Med*. 2017;47(1):113–26.
10. Whitehead M. Physical literacy: philosophical considerations in relation to developing a sense of self, universality and propositional knowledge. *Sport Ethics Philos*. 2007;1(3):281–98.
11. Robinson DB, Randall L. Marking physical literacy or missing the mark on physical literacy? A conceptual critique of Canada’s physical literacy assessment instruments. *Meas Phys Educ Exerc Sci*. 2017;21(1):40–55.
12. Dudley D, Cairney J, Wainwright N, Kriellaars D, Mitchell D. Critical considerations for physical literacy policy in public health, recreation, sport, and education agencies. *Quest*. 2017:1–17.

13. Canadian Sport for Life. *Physical Literacy Assessment for Youth*. Victoria (BC): Canadian Sport Institute; 2013.
14. Longmuir PE, Boyer C, Lloyd M, et al. The Canadian Assessment of Physical Literacy: methods for children in grades 4 to 6 (8 to 12 years). *BMC Public Health*. 2015;15(1):767.
15. Physical & Health Education Canada. *Passport for Life*. Ottawa (ON): PHE Canada; 2013.
16. Ontario Ministry of Education. Ontario Health and Physical Education Curriculum, Grades 1 to 8 [Internet]. Toronto, ON. 2015. Available from: <http://www.edu.gov.on.ca/eng/curriculum/elementary/health1to8.pdf>.
17. Manitoba Education and Training. Manitoba Physical Education/Health Education Curriculum [Internet]. Winnipeg, MB. Available from: <http://www.edu.gov.mb.ca/k12/cur/physlth/curriculum.html>.
18. Canadian Sport for Life. *Long-Term Athlete Development Resource Paper 2.1*. Victoria (BC): Canadian Sport Institute; 2016. pp. 23–5.
19. Henderson S, Sugden D, Barnett A. *Movement Assessment Battery for Children*. 2nd ed. Pearson Education: London (UK); 2007.
20. Bruininks R, Bruininks B. *Bruininks–Oseretsky Test of Motor Proficiency*. 2nd ed. Minneapolis (MN): AGS Publishing; 2005.
21. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. New York (NY): Oxford University Press; 2014. pp. 38–64.
22. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. 1999;6(1):1–55.
23. Yu C, Muthen B. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. In: *Proceedings of the Annual Meeting of the American Educational Research Association*. 2002 Apr; New Orleans (LA). pp. 1–5.
24. 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(2):162–70.
25. Thomas JR, French KE. Gender differences across age in motor performance: a meta-analysis. *Psychol Bull*. 1985;98(2):260–82.
26. Payne V, Isaacs L. *Human Motor Development: A Lifespan Approach*. 8th ed. New York (NY): McGraw-Hill; 2011. pp. 352–416.
27. Koivula N. Ratings of gender appropriateness of sports participation: effects of gender-based schematic processing. *Sex Roles*. 1995;33(7–8):543–57.
28. Koivula N. Perceived characteristics of sports categorized as gender-neutral, feminine and masculine. *J Sport Behav*. 2001;24(4):377–93.
29. Hardin M, Greer J. The influence of gender-role socialization, media use and sports participation on perceptions of gender-appropriateness of sports participation. *J Sport Behav*. 2009;32(2):207–26.
30. Barnett LM, Hardy LL, Brian AS, Robertson S. The development and validation of a golf swing and putt skill assessment for children. *J Sports Sci Med*. 2015;14(1):147.
31. Staples KL, Reid G. Fundamental movement skills and autism spectrum disorders. *J Autism Dev Disord*. 2010;40(2):209–17.
32. Zask A, Barnett LM, Rose L, et al. Three year follow-up of an early childhood intervention: is movement skill sustained? *Int J Behav Nutr Phys Act*. 2012;9(1):127.
33. Holtfelder B, Schott N. Relationship of fundamental movement skills and physical activity in children and adolescents: a systematic review. *Psychol Sport Exerc*. 2014;15(4):382–91.
34. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods*. 1999;4(1):84–99.