E EA CH A

CrossMark

Refining the Canadian Assessment of Physical Literacy based on theory and factor analyses

Katie E. Gunnell[1,2*], Patricia E. Longmuir[1], Joel D. Barnes[1], Kevin Belanger[1] and Mark S. Tremblay[1]

A

Background: The Canadian Assessment of Physical Literacy (CAPL) is a 25-indicator assessment tool comprising four domains of physical literacy: (1) Physical Competence, (2) Daily Behaviour, (3) Motivation and Confidence, and (4) Knowledge and Understanding. The purpose of this study was to re-examine the factor structure of CAPL scores and the relative weight of each domain for an overall physical literacy factor. Our goal was to maximize content representation, and reduce construct irrelevant variance and participant burden, to inform the development of CAPL-2 (a revised, shorter, and theoretically stronger version of CAPL).

Methods: Canadian children (n = 10,034; M_{age} = 10.6, SD = 1.2; 50.1% girls) completed CAPL testing at one time point. Confirmatory factor analysis was used.

Results: Based on weak factor loadings (s < 0.32) and conceptual alignment, we removed body mass index,

B

Over the past decade, researchers, practitioners, and teachers have become interested in the concept of physical literacy given its relevance to healthy active living, physical education curricula, policy, public health, sport, and active recreation [1–3]. Although global consensus across researchers and practitioners on the definition of physical literacy has yet to be reached [4], in 2015 several Canadian organizations collectively adopted and recognized the definition set forth by the International Physical Literacy Association [5–7]. In Canada's Physical Literacy Consensus Statement (Canadian Consensus Statement) [5], physical literacy is defined as the "motivation, confidence, physical competence, knowledge and understanding to value and take responsibility for engagement in physical activities for life" [5].

Alongside the rapid proliferation of interest and research on physical literacy grew the need to develop an assessment tool that could be used to derive valid and reliable physical literacy scores [2, 3, 8]. Although several instruments have been created to measure physical literacy [2], the Canadian Assessment of Physical Literacy (CAPL) is the only assessment of children's physical literacy that has undergone extensive peer-reviewed and published validation efforts, including assessments of feasibility, validity, and reliability [8, 9]. The purpose of this study was to update the validity evidence for CAPL scores, with the goal of reducing participant burden while also maximizing validity evidence based on factor structure and content representation aligned with recent advances in physical literacy research and theory.

Canadian assessment of physical literacy (CAPL)

The CAPL was developed to meet the demand for an assessment tool that could be used to produce valid and reliable scores that were representative of children's progress on their physical literacy journey [9]. The creation of the CAPL involved consultation with practitioners (e.g., physical education teachers) and researchers, an extensive review of Canadian school physical education curricula, the identification of existing assessments, and the creation of novel assessments when no others existed (see [10 of physical fitness were included within CAPL given that Whitehead [16] has argued that physical literacy is a journey that requires the capacity to be physically competent. CAPL developers have argued that the measurement of fitness can be used to indicate an individual's capacity to sustain physical activity for life [17]. Similarly, the creators of CAPL interpret the Canadian Consensus Statement definition to mean that when people value and take responsibility for engaging in physical activity, they will demonstrate this by being physically active. From this perspective, the creators of CAPL believe the definition of physical literacy includes physical activity behaviour as part of physical literacy. This perspective is consistent with the model proposed by Robinson and Randall [2] and the elements of physical literacy outlined in the Canadian Consensus Statement [5, 6].

CAPL indicators for representing their respective domains. For example, standardized body mass index (BMI) z scores, total physical activity scores derived from pedometer step counts, and screen time scores were weak indicators (λs < 0.27) of their respective CAPL domains [10].

In the present study, using a much larger dataset, we re-examined validity evidence of scores from the full CAPL protocol based on factor structure and content representation alongside recent advances in the conceptualization of physical literacy, and relying on scientific and theoretical advances in physical literacy research. A second purpose was to reassess the relative weight of each domain for creating an overall physical literacy score. Our overall goal was to maximize content representation and validity evidence while reducing construct irrelevant variance and participant burden. The results from this study will set the stage for the development of CAPL-2 – a condensed and more theoretically aligned version of CAPL.

It is important to recognize that in this study we were not examining the validity of each indicator (e.g., grip strength) but, rather, examining how each indicator coalesces to demarcate a CAPL domain. Therefore, although we will present evidence for a reduced model of the CAPL, we do not wish to imply that the indicators removed were in some way invalid, but rather that they do not configure optimally, theoretically, statistically, or logistically in combination with the other indicators to create a particular CAPL domain. It is also important to note that initially we took a confirmatory approach to examine the a priori hypothesized model of CAPL based on past theory and empirical evidence [8, 9]. When the model did not fit the data, a more exploratory approach through the lens of confirmatory factor analysis was used; however, to avoid the negative consequences of data-driven specification searches [18], extensive discussion of substantive theory and alignment preceded data-driven modifications. Exploratory modifications were not made unless they were theoretically anchored and were supported by all co-authors. Further, all modifications were made with suggestions for cross-validation in future research.

M

Participants and procedures

A complete description of the participants and procedures for these data are presented by Tremblay and colleagues in this issue [19]. Briefly, Canadian children (n = 10,034; 50.1% girls) ranging in age from 8.0 to 12.9 (M_{age} = 10.6, SD = 1.2 years) completed CAPL testing. This project was approved by the Children's Hospital of Eastern Ontario Research Ethics Board (Ottawa, Ontario; coordinating centre) and research ethics boards at each site. Parents or legal guardians provided written informed consent and children provided verbal assent.

CAPL measures

The CAPL is comprised of standardized assessments with evidence of score reliability and validity in children aged 8 to 12 years [8, 14, 15, 20–25]. A detailed description of each assessment protocol can be found at www.capl-ecsfp.ca.

Physical competence

The Physical Competence domain was assessed with 7 indicators. (1) Movement skills were assessed with the Canadian Agility and Movement Skill Assessment (CAMSA; [15, 23]. The CAMSA can be used to assess children's fundamental (e.g., sliding, catching, jumping, throwing, skipping, kicking, hopping), complex (e.g., acceleration and deceleration, rhythmic movement, hand-eye coordination),nippio8M3403-1.61(ion)8(rh00510g.)-370.39999

Mc

(coantch)-11.69999c1(be)0878(wit719108.59997558(tbe)0108.599975gu weekend day (a) watching television or (b) playing video games and/or using computers in recreational time with response options of 0 (none) to 6 (5 h or more). An included a combination of categorical and continuous indicators, the mean- and variance-adjusted weighted least square (WLSMV) estimator was used.

Each model was evaluated and modified using a combination of theory, statistical criteria (described below), and group consensus amongst all authors until an acceptable model was retained. For the Motivation and Confidence domain, a supplemental model was examined to determine if the indicator of preferred leisure activities fit under the domain of Motivation and Confidence (see "CAPL measures" section for justification of this supplemental analysis). Next, the final models from each individual domain's confirmatory factor analysis were combined to examine a four-factor correlated measurement model using the WLSMV estimator attributable to the combination of continuous and categorical data. Finally, a higher-order model was tested to examine if physical literacy loaded onto each of the four domains that were measured by their respective indicators, and to determine the relative magnitude of each loading to inform the configuration of CAPL domains. To ensure missing data were not affecting the results, we conducted a sensitivity analysis to examine if this higher-order model yielded a similar pattern of results when using only participants who had complete data compared to including participants with incomplete data.

Evaluating model fit

Model fit for each factor analysis was examined alongside theoretical considerations and statistical guidelines, group consensus, and, in some cases, feedback received from CAPL experiences. A combination of goodness-of-fit statistics was used to determine statistical data-model fit, while recognizing that p

1 Confirmatory factor analysis results of Physical Competence

	Model 1			Model 2			Model 3		
		SE	R^2		SE	R^2		SE	R^2
PACER	0.283*	0.012	0.080	0.799*	0.010	0.639	0.800*	0.011	0.640
Plank	0.273*	0.011*	0.075	0.564*	0.010	0.318	0.551*	0.010	0.303
CAMSA	0.094*	0.015	0.009	0.613*	0.010	0.375	0.602*	0.010	0.363
Grip strength	−0.420*	0.010	0.177	0.266*	0.014	0.071	–	–	–
Sit and reach	0.142*	0.010	0.020	0.161*	0.013	0.026	–	–	–
BMI z (reverse coded)	0.802*	0.012	0.643	0.318*	0.012	0.0101	–	–	–
Waist circumference (reverse coded)	0.999*	0.012	0.998	0.285*	0.012	0.081	–	–	–

*$p < 0.01$, = factor loadings, –- = item was not included in the model

model convergence. Model 1 was just identified and therefore fit statistics could not be obtained. Inspection of the factor loadings indicated that self-reported screen time had the lowest factor loading ($\lambda = 0.35$; see Table 2). Although screen time and sedentary behaviours are conceptualized under the movement continuum [32], we reasoned that screen time might not belong conceptually in a measurement model of physical literacy given recent evidence showing that physical activity and sedentary behaviour are separate and weakly correlated movements [33, 34]. Therefore, a decision was made to remove self-report screen time from the model. Finally, although the factor loading of pedometer step counts was relatively weak ($\lambda = 0.40$), pedometers are considered to be a more direct indicator of physical activity compared to self-report physical activity [35]. As such, it was retained in the model on conceptual and content representation grounds. Model 2, removing self-report screen time from daily behaviour, could not be estimated with only two indicators. Therefore, the factor loadings of the Daily Behaviour domain were further examined in the full measurement model.

Motivation and confidence

Model 1 of the Motivation and Confidence domain provided an unacceptable fit to the data ($MLR\chi^2_{(5)} = 1529.61$, $p < 0.001$, $CFI = 0.876$, $RMSEA = 0.176$, 90% CI [0.168, 0.183]; see Table 3). Modification indices (Modification Index = 1265.62) suggested that there was a large error covariance between the indicators of "activity compared to others" and "skill compared to others". We reasoned that this error covariance could be attributable to the similar wording and response options of these questions. Model 2, with an error covariance between these two indicators, provided an excellent fit to the data ($MLR\chi^2_{(4)} = 365.42$, $p < 0.001$, $CFI = 0.971$, $RMSEA = 0.096$, 90% CI [0.087, 0.104] $r_{\text{activity compared to others and skill compared to others}} = 0.42$, $p < 0.01$; see Table 3). Despite excellent model fit, we removed "activity compared to others", for two reasons. First, the two items with the error covariance were very similar and might therefore cause redundancy in the model (i.e., each is not adding unique construct relevant variance). Second, the indicator "activity compared to others" had the lowest factor loading compared to the indicator "skill compared to others". The model was re-estimated excluding "activity compared to others." Model 3 for Motivation and Confidence provided an excellent fit to the data ($MLR\chi^2_{(2)} = 188.94$, $p < 0.001$, $CFI = 0.979$, $RMSEA = 0.097$, 90% CI [0.086, 0.109]; see Table 3) and was retained as the final model.

At this point, a supplemental analysis was run to examine if specifying activity preferences as an indicator of Motivation and Confidence provided a good fit. In this analysis, Model 3 of motivation and confidence was re-estimated, including the categorical indicator of activity preferences and using the WLSMV estimator given the combination of categorical and continuous indicators.

Compared to Model 3 of motivation and confidence, including the activity preference indicator degraded model fit (WLSMV$\chi^2_{(5)} = 509.66$, $p < 0.001$, CFI = 0.936, RMSEA = 0.101, 90% CI [0.094, 0.109]). Although the indicator of activity preferences loaded relatively strongly on motivation and confidence ($\lambda = 0.576$, $p < 0.01$), a decision was made to not include this indicator because it decreased overall model fit of the domain and because it is a dichotomous indicator that may not provide as much content representation compared to the current indicators of motivation and confidence.

Knowledge and understanding

Model 1 with all original indicators of Knowledge and Understanding provided an unacceptable model fit (WLSMV$\chi^2_{(35)} = 627.92$, $p < 0.001$, CFI = 0.881, RMSEA = 0.041, 90% CI [0.039, 0.044]; see Table 4). In Model 2, the indicators of safety, activity preferences, and screen time guidelines were removed o response. In fact, 63% of children answered this question correctly, suggesting that the response options should be altered to provide better discrimination. Based on this informal observation, along with considerations of maximizing content representation for the Knowledge and Understanding domain in light of the number of indicators already removed, a decision was made to retain the physical activity guideline indicator given its conceptual relevance. Overall, decisions around retaining weak indicators in the Knowledge and Understanding domain (i.e., “physical activity guidelines” and “improve sport skills”) were made with an eye toward optimizing construct representation while recognizing that these items require refinement if they are to be included in CAPL-2. Model 3 provided a good fit to the data (WLSMV$\chi^2_{(5)}$ = 97.39, p

Higher-order model of physical literacy

In the final model tested, we examined whether an overall physical literacy latent variable accounted for the correlations between the four domains. Using the final model from the four correlated domains model above, we found that the higher-order model had a good fit to the data (WLSMV$\chi^2_{(72)} = 1827.18$, $p < 0.001$, CFI = 0.919, RMSEA = 0.049, 90% CI [0.047, 0.051]; see Fig. 2). Daily Behaviour, Motivation and Confidence, and Physical Competence had the strongest factor loadings from physical literacy. Knowledge and Understanding had a significant, albeit weak, factor loading (see Fig. 2). Therefore, for the next version of CAPL, it is recommended that the domains be re-weighted such that Physical Competence, Daily Behaviour, and Motivation and Confidence have stronger weight (30 points each) than Knowledge and Understanding (10 points). The revised model suggested for CAPL-2, based on the good factor structure suggesting four intercorrelated domains of physical literacy, is shown in Fig. 1d.

An additional analysis using listwise deletion to examine the final higher-order model with only participants who had complete data ($n = 5073$) yielded a similar pattern of factor loadings on the pedometer step counts scores and self-report physical activity indicators (specific results from this supplemental analysis are available from the corresponding author upon request).

D[illegible]

We examined the 25-indicator CAPL using factor analytic techniques to examine the validity of evidence based on factor structure and to determine the relative weighting of each domain of physical literacy. Using confirmatory factor analyses to test the a priori specified CAPL model and exploratory post-hoc modifications based on theory, group consensus, and statistical criteria, we found support for a revised and conceptually concise 14-indicator version of CAPL that maximized content representation while redu-

CAPL administrators through informal feedback, the response option for the physical activity guideline question may have been too easy given that the correct response was the highest value listed. This response option is problematic because it is easy to guess and might have been cued by the self-report physical activity indicator, which includes “60 min” in the instructional stem. In other words, it is possible that the two questions are sharing variance given that children can link the responses of one to the stem of the other. Future research is needed to determine if altering the response options for the knowledge of physical activity guidelines can reduce this problematic cross-loading. Finally, it is worth noting that we did not remove knowledge of physical activity guidelines as an indicator of the Knowledge and Understanding domain and place it as an indicator of Daily Behaviour. Our decision to retain it as an indicator of knowledge and understanding and allow for the cross-loading was based on the conceptual content of the indicator. Because the indicator asks children to report how long kids should be active, we did not feel it was a good indicator of how long the children themselves were actually active.

Five out of 10 indicators from the Knowledge and Understanding domain were removed. Conceptually, many original indicators that have not undergone previous validation do not align with current physical literacy research. For example, knowledge of what it means to be healthy could be an indicator of health literacy rather than physical literacy. Removing self-report screen time from the Daily Behaviour domain and knowledge of screen time guidelines from the Knowledge and Understanding domain can be justified for similar reasons; they may not be conceptually linked to physical literacy given in physical activity situates physical activity as an inherent component of physical literacy. Nonetheless, a strength of the CAPL is that researchers are free to use (or not use) domains that are of interest to them, given their theoretical perspective. As such, it is possible to use the CAPL and omit the Daily Behaviour domain to assess physical literacy, as described in the CAPL Manual (available at www.capl-ecsfp.ca).

A key finding of this study, when taken in the context of past versions, conceptualizations, and operationalizations of the CAPL, is that physical literacy cannot be reduced only to fitness or motor skill assessments. Indeed, original conceptualizations placed strong emphasis on indicators of physical fitness; yet over the past decade, with emerging validation and physical literacy research, it has become apparent that physical competence encompasses more than fitness and body composition and that motivation and confidence are equally important for physical literacy.

Strengths, limitations and future directions

A major strength of this study was the large sample size, which included children from 11 regions across Canada, enabling the assessment of all CAPL indicators. Notwithstanding these strengths, limitations of this study should be acknowledged. There was a large amount of missing data on the pedometer daily step count scores, and there were gender differences in the amount of missing data for pedometers. Although modern procedures were used to handle missing data, it is not known to what extent the missing data influenced the final conclusions. Nonetheless, sensitivity analysis of the final higher-order model without missing data (i.e., using listwise deletion $n = 5073$) revealed a similar pattern and magnitude of results to those we report herein. Second, although our decisions to remove or retain indicators were made based on statistical criteria, theory, and considerations of content representation, many of these decisions were ultimately subjective in nature. It is possible that alternative models would have fit the data well. Moreover, although we used confirmatory factor anadata acquisition online, prepared the data for analysis, and contributed to the data analyses and interpretation of results as well as manuscript preparation. KB oversaw the acquisition of data, and contributed to interpretation of the results and manuscript preparation. MST developed the original CAPL, led the acquisition of data, and contributed to the interpretation of the results. All authors read, revised, and approved the final manuscript.

Ethics approval and consent to participate

Ethics approval was obtained from: Antigonish - St. Francis University Research Ethics Board and the Strait Regional School Board; Calgary - Mount Royal

33. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms