Development of ELIP (*Évaluation de La LIttératie Physique*) to assess physical literacy for emerging adults: a challenge embraced through Delphi method and cognitive interview process.

Words: 7372

Abstract

Following increased interest in physical literacy (PL), development of appropriate tools for assessment has become an important next step for its operationalization. To forward the development of such tools, the objective of this study was to build the foundations of the Évaluation de La LIttératie Physique (ELIP): designed to help reduce existing tensions in approaches to PL assessment resulting in a low uptake in applied settings. We followed two steps: (1) the development of the first version of ELIP by deploying a Delphi method (*n*=30); and (2) the modification of items through cognitive interviews with emerging adults (*n*=32). The expert consensus highlighted four dimensions of PL to be assessed – physical; affective; cognitive; and social – with new perspectives, including a preference for broad motor tests over fitness. Results offer new insights into the assessment of emerging adults' PL but ELIP still requires further work concerning validity, reliability, and sensitivity.

Keywords: Assessment, Cognitive interview, Delphi method, Physical literacy

Recently, the concept of Physical Literacy (PL) (Whitehead, 2001, 2010) has gained increased attention in international literature (Young et al., 2020). It has become a key focus on the issue of the promotion of physical activity (PA) in different contexts (e.g., health, sport, physical education, and recreation), because it helps to shift the argument from simply being active, to accruing holistic benefits through physical movement (Whitehead, 2001). PL is derived from the wider generic term, literacy, understood as the essential part of an individual's education and an essential component for participation in society (UNESCO, 2005), and so PL is positioned as an individual's foundation for a healthy physical life (Cairney et al., 2019a).

The concept of Physical Literacy: one concept and several definitions

Whilst many discrepancies about the definitions remain in the literature (Edwards et al., 2017; Martins et al., 2020), most PL experts reject a Cartesian distinction between mind and body and rather promote the idea of richly holistic and embodied learning in order to promote a beneficial and enriching lifelong relationship with movement and PA (Pot et al., 2018; Whitehead, 2007). The International Physical Literacy Association (IPLA) has defined PL specifically from Whitehead's perspective: "*Physical Literacy can be described as the motivation, confidence, physical competence, knowledge and understanding to value and take responsibility for engagement in physical activities for life*" (IPLA, 2017). Despite its international popularity (Edwards et al., 2017) – noting that popularity is not necessarily the same as veracity – additional definitions suggest other perspectives, and may provide additional insight into what PL is (e.g., Keegan et al., 2019). In a recent European consensus statement, PL appeared as "*the skills and attributes individuals demonstrate through physical activity and movement across their life course. It can be understood as a process and as an outcome that individuals pursue through an interaction of their physical, emotional, social and cognitive learning*" (Physical Literacy For Life, 2021). This definition, close to the

Australian conception (Keegan et al., 2019), highlights the necessary holistic development of people through movement and suggests that different outcomes, or milestones, can be set across the lifetime through measurement tools.

The philosophical challenge of assessing Physical Literacy in the emerging adult period

Due to concerns over both PL's pedagogical positioning and epistemology, its assessment is increasingly questioned by researchers (Chen, 2020a; Edwards et al., 2018; Green et al., 2018). This reflection has become one of the three prominent themes in the growing literature related to PL: assumptions of the concept and its educative role, sports development, and PL assessment (Lundvall, 2015). To date, it has been challenging to reach agreement on measurement tools that appropriately accommodates the epistemological foundations (Liu & Chen, 2020; Robinson & Randall, 2017; Young et al., 2020) while adopting a pragmatic form.

Partly due to its development from multiple philosophical perspectives – especially monism, existentialism and phenomenology (Pot et al., 2018) –the assessment of PL remains a sensitive issue among experts and thus it is an ongoing challenge (Chen, 2020a; Edwards et al., 2018; Longmuir & Tremblay, 2016; Robinson & Randall, 2017; Whitehead, 2010). A recent literature review reported different PL assessments (Edwards et al., 2018) and that the overwhelming majority are detached from the holistic perspective that PL is intended to promote, frequently focusing on the motor skills. Nevertheless, different tools which match the holistic perspective of PL assessment have been proposed recently - PPLA-Q (Mota et al., 2021); PL-C Quest (Barnett et al., 2020); CAEPL (Chen et al., 2020b)- and are added to well-known tools like the CAPL (Longmuir et al., 2015), PPLI (Sum et al., 2016), PFL (Lodewyk, 2019), PLAY tools (Caldwell et al., 2021), or Pre-PLAY (Cairney et al., 2018) - but there remain opportunities to advance measurement capability for PL: both in terms of broadening

the population assessed and respecting the philosophical foundations of PL (Pot et al., 2018; Whitehead, 2010).

PL assessment has largely focused on school-age children. However, PL is a concept that must be understood throughout the life course (Cairney et al., 2019b; Keegan et al., 2019), and so its assessment also needs to be extended to other age groups according to their characteristics. The singular characteristics of emerging adults, such as their new relationship with them body and to other people (Berndt & Savin-Williams, 1993), needs to be considering in a new measurement tool. In respect to this study, 'emerging adults', defined as the time from the end of adolescence to young-adulthood responsibilities, have to be a target population for PL assessment tools (Edwards et al., 2018; Longmuir & Tremblay, 2016). This missing assessment period (16-21 years old), without a corresponding tool, is associated with an important life transition for individuals (Arnett, 2000). It can be considered as a key period during the PL journey where French individuals are, for the last time, engaged in compulsory physical education and where the most active people are still engaged in organized PA (Muller, 2018). Interest regarding the PL of this target population is emerging (Kwan et al., 2019), but to date specific measurement and assessment instruments are missing.

The way in which the level of PL development is scored is a controversial issue (Chen, 2020a). At present, there is a discrepancy between the comprehensive Whiteheadian philosophical basis of PL and the tools for measuring it. The pre-existing tools seem to fail to fully match the scorers' processes of PL with the philosophical underpinnings of the concept (Chen, 2020a; Robinson & Randall, 2017). The linear and simple arithmetic approaches, consisting of summative scores by dimension (e.g., PPLI or PLAY) or to attribute less importance to one dimension than to others (e.g., CAPL), are questionable in terms of monist aspect of PL (Chen, 2020a). The monist perspective, considering the individual as a whole in which each component interacts with the others with equal importance, could be further

enhanced. Indeed, the complex nature of PL is not fully reflected by these approaches since human behavior does not follow a linear logic and should be apprehended through complex systems (Heino et al., 2021). Scoring procedures can be improved by considering the idea of interdependence and a similar significance between each of the PL dimensions, which corresponds to the initial nature of the PL concept (Dudley, 2018). The added value of PL regarding physical lifestyles is therefore still difficult to study empirically, although early evidence suggests PL does represent this superior construct going beyond the juxtaposition of dimensions (Cairney et al., 2019a). The identification of the constitutive PL components is also a matter of controversy in the literature (see Robinson & Randall, 2017; Tremblay & Longmuir, 2017) and needs to be further questioned. A new PL framework (defining statements, domains, elements and guidelines for development) (Keegan et al., 2019), philosophical perspectives (Rudd et al., 2020), and recent work on the fundamental/foundational skill determinants of a sustainable physical life (Hulteen et al., 2018) have all advanced the literature while also reframing the debate about how best we may measure PL as well as in populations other than children.

Likewise, some tools diverged from the existentialist philosophy advocated by Whitehead. For example, by imposing a restrictive sports-oriented vision of PA (e.g., PPLI) or not really challenging the participant to adapt freely to a complex task (e.g., CAPL), these tools may not have garnered the breadth of information necessary to respect the indivisible couple individual/environment(Whitehead, 2001). Also, to capture the most valuable information in this perspective, a mixed tool combining questionnaires and motor tests could be necessary. Despite the highly pragmatic nature of questionnaires, it could be considered reductive to focus only on questionnaire responses. Further to this, early studies exposed differences between perceived and actual level of PL or physical competences (Barnett et al., 2015; Li et al., 2021).

Finally, the phenomenological perspective is also a sensitive question in an empirical PL evaluation issue. This approach assumes that each individual and his or her relationship to the world is unique (Merleau-Ponty, 1968) and therefore, this conception is opposed to a normative assessment, but instead privileges an ipsative evaluation (in comparison only with one's own previous measure) or 'charting' (Whitehead, 2019). Despite the debates, even the most vocal opposition to PL assessment still concede the importance of this issue for the development of the concept and the promotion of PA in life (Robinson & Randall, 2017; Whitehead, 2019). Thus, we believe that one current challenge is to build a tool which, can provide the necessary information for the empirical study of the concept and to inform researchers as well as practitioners of individual's evolution at the different milestones of his/her PL journey. The existence and insights offered by such a tool will help to design educational programs and to identified learners' development opportunities in order to foster individuals PL development, while being a positive step in PL journey for each participant. The design of a such measurement tool will also support the theoretical foundation of PL, including identifying whether certain typical PL profiles are favorable to a healthy commitment in PA through emerging adult life.

A methodological challenge to produce a valid, reliable, and practical tool

Thus, to enable the evaluation of interventions and the informing of day-to-day practices of those invested in PL, there is a need to design assessment specific tools for the target population. Nevertheless, developing a new measurement tool for PL is a scientific and methodological challenge both in terms of validity and reliability (Gunnell et al., 2014; Mokkink et al., 2010).

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014, p.11) claimed that validity is "*the degree to which evidence and theory support the*

interpretations of test scores for proposed uses". According to the seminal work of Messick (1995), six aspects are crucial to establish this construct validity: content, substantive, structural, generalizability, external, and consequential aspects. Validation practices by researchers in sport sciences and sport psychology have been variously critiqued before, challenging the robustness of approaches in this domain (Gunnell et al., 2014; Zhu, 2012). Importantly, achieving the challenge of building an appropriate PL measurement tool would help promote the concept in the scientific and professional community. The quality of a tool is also verified by its reliability (Mokkink et al., 2010). The development of a new PL measurement tool requires a rigorous step-by-step approach that needs to be conducted carefully and systematically.

To date, there is a need for new tools, especially for emerging adults, to overcome the limits of the previous ones to further study and operationalize the PL concept respecting is philosophical basis. Hence, to contribute to the ongoing development of PL assessments, we set out to develop a new tool that: (1) adequately capture main philosophical underpinnings of PL; (2) set milestones for testing PL that goes beyond children; (3) question the combination of subjective and objective assessments; and (4) respect the main steps of a robust validation of a measurement tool. We therefore undertook two studies to start designing this tool to integrate two different and complementary points of view: PL experts and emerging adults. Our research question was simply how we can robustly assess the PL of emerging adults?

Methods

We conducted two studies, consecutively, to develop a tool to assess the level of PL development in emerging adults (Figure 1). The first study deployed a Delphi method process (Powell, 2003), to reach a consensus about the structure and the design of the assessment tool (i.e., face and content validity). The second study used a cognitive interview process to

ascertain the feasibility and adequacy of the question's meaning and interpretation of the respondents (Willis, 2005).This additional insight is rarely included in the development of such assessments, particularly in **physical activity and sport sciences**, but is necessary to increase reliability and validity (Dietrich & Ehrlenspiel, 2010). All the steps were in consistence with the COSMIN recommendations (Mokkink et al., 2010).

[Figure 1 near here].

Study 1

Participants - Recruitment of expert panel. Our selection process was informed by considerations for (1) recruiting experts who recently published articles related to PL in journals indexed in PubMed database; (2) recruiting experts of different nationalities; and (3) gathering different points-of-view through expertise from different scientific fields. Among 31 experts contacted, 20 responded favorably (66.67% positive responses). We used snowball sampling by inviting experts to propose other experts (Hanson et al., 2020). This resulted in the identification of 15 other individuals not originally identified in the database search. The analysis of their professional background made it possible to include them in the panel of experts and 10 responded favorably (66.7% positive responses). Professional backgrounds included: (1) published papers on PL; (2) published papers on engagement in PA; and (3) involvement in research or professional group on these subjects.

A five-point scale (1= beginner / 5= expert) was completed by each expert to self-assess their PL expertise; the mean response was 4.63±0.49. The summary description of the participants is shown in Table 1. To thoroughly involve the experts, we offered panel members the opportunity to become co-authors on the final publication generated by the study and included verification-questions in the questionnaires (see section below). In the following reporting and during the process, all data have been anonymized to minimize bias.

[Table 1 near here].

Delphi process to develop ELIP structure. According to Chen (2020a) and Edwards et al. (2017), the debate on how to assess the level of PL needs to be raised among all experts. The goal of the Delphi process therefore was used to reach a consensus on a specific research issue after successive rounds of discussions. It is defined as "*an iterative process, designed to combine expert opinion, in order to arrive at a group consensus*" (Keegan et al., 2019, p. 3) and is recommended when discordance surrounds a topic (Powell, 2003).

Delphi method's design. We implemented four iterative rounds of Delphi method through the LimeSurvey platform (https://www.limesurvey.org/fr/) from May to October 2020. The successive rounds aimed to: (1) examine the constitutive dimensions of the tool; (2) examine the constitutive domains of each dimension; (3) examine tests to assess the accepted domains; and (4) design a filter to transform the items to be in line with the epistemological foundations of the PL concept. In each Delphi round, the results of the previous round were presented. Experts not responding despite two reminders were removed from the concerned round but were invited to participate in the following rounds. (n<8). All the selected participants consider themselves as PL experts and anchored in different PL perspectives. Thus, no definition was imposed on the panel to design the ELIP, and therefore the result is based on the elements that reached consensus.

Questionnaire design. The literature was screened to identify the potentially relevant dimensions to be included in the ELIP. The first round involved capturing the relevance and comments about the first selection of PL dimensions (Supplementary File 1). The second round consisted of capturing the relevance and comments about a first selection of PL domains for each accepted dimension (Supplementary File 2). Domains are understood as the constituent elements of the dimensions. Based on the results of the previous two steps, the

219 literature was analyzed to identify and/or design relevant tests to assess each accepted
220 domain. For this selection, the research team questioned the content and face validity and
221 examined the relevance and sensitivity for a French emerging adult population of each
222 different tools identified. The relevance of these tests was verified in round 3 (Supplementary
223 File 3). To transform these tests as close as possible to the PL concept, we collected the
224 comments from Round 3 and reformulated them to create 'filters' (i.e., sentences that we use
225 to transform the test items). In a last round, experts were questioned about the relevance of the
226 filters (e.g., "*How much do you agree with this idea? 'The items of social dimension should*
227 *be causally related to physical activity and not remain general'*"). Finally, each accepted item
228 of the selected tests was transformed through the filters until a consensus between the initial
229 team (GJ, DT, PF, SC). At the end of this step, we had a set of items to submit to experts for
230 content validation.
231 For each round, experts responded to an online questionnaire and were invited to rate
232 the relevance of the different propositions thanks to a Likert Scale from 1 to 5 ("*not relevant*
233 *at all*" to "*very relevant*"). The use of open-ended questions helped to obtain richer data
234 (Powell, 2003), and experts were free to comment on the different domains proposed and to
235 make suggestions.
236
237 ***Consensus requirements.*** Each round ended only when a consensus was reached.
238 Usually, an item is retained in Delphi method if it is accepted by most of the group's experts:
239 between 55% and 100% (Powell, 2003). Here, a limit of 66% was tolerated (Powell, 2003)
240 considering the divergence existing in terms of definitions (L. C. Edwards et al., 2017;
241 Shearer et al., 2018), as well as the sensitivity of the issue of PL assessment (Chen, 2020a; L.
242 C. Edwards et al., 2018). Otherwise, the item is either requested in sub-round if agreement
243 between 61% and 65% or discarded if less than 61%. The comments submitted by the experts

were collected and those with at least two converging comments were considered for a next step. Two comments converge when they propose, in a different way or not, a common element (e.g., "*Beliefs should not be in this dimension but in the cognitive dimension"* and "*beliefs are primarily cognitive*").

Content Validity. Each item reformulated through the filters and instruction was then subjected to content validation by the same 30 experts. "*It is the degree to which the elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose"* (Haynes et al., 1995, p. 14). Expert's judgements on the relevance, representativeness, and clarity were captured through a 1 to 5 Likert scale. We followed the recommendations of Lawshe (1975) to consider the acceptance standard of Content Validity Ratio (CVR) ($n \leqslant 25$, CVR $\geqslant 0.37$). At the end of this step, we had a set of items to be specified by the Cognitive Interview (CI) method with emerging adults.

Results

Delphi Method

In total, we invited 46 experts to participate, 31 from literature reviews and 15 via snowball sampling process. We received 30 positive responses (65.2%). Our results are well within the minimum number of 15 experts (Mokkink et al., 2010).

Round 1. The results of the first two rounds are presented in Table 2. The 23 experts (76.6%) who completed the first round strongly accepted to integrate in the ELIP the three following domains: physical (100%), affective (100%), cognitive (95,6%) and slightly less the social dimension (86,9%). According to the experts, all four dimensions must be integrated into the ELIP. Within this study, experts privilege objective rather than subjective tests for the physical dimension and any proposals for other dimensions emerged from the comments left

by the experts. Despite the choice to focus on objective tests, experts chose to evaluate perceived aquatic competencies rather than the objective ones (73.9%).

Round 2. The questionnaire was completed by 23 experts (76.6%). In the first sub-round, one affective domain, four social domains and three cognitive domains were directly discarded (agreement<61%). Beliefs (affective), cardiorespiratory capacity (physical) and strength (physical) were re-questioned in a second sub-round (61%>agreement<66%). In this second sub-round, physical fitness (including strength and cardiorespiratory capacity) was discarded, and the belief domain was moved onto the cognitive dimension (Table 2).

[Table 2 near here].

Round 3. The results of the third round are presented in Table 3. Twenty-four experts (80%) completed this round. Fourteen tests were submitted to the experts for their opinion. Eight tests were directly accepted (agreement>66%) and one was directly discarded (agreement<61%). Three tests were questioned in a second sub-round according to the comments from the experts and were accepted with slight modifications (e.g., *focusing only on one part of the questionnaire*).

[Table 3 near here].

Round 4. The results of the fourth round are presented in Table 4. We reworded all comments that appeared at least twice in Round 3 as filters. These filters transformed the accepted test items to fit the epistemological foundation of PL. Sixteen (16) filters were drafted and proposed for experts' validation (Table 4). Twenty-five (25) experts (83.3%) accepted thirteen (13) filters (agreement>66%). None were challenged for a second sub-round.

[Table 4 near here].

Content validity. Twenty-five experts (80.6%) accepted 88 items (CVR$\geqslant$0.37). In addition, the completeness of the tool's dimensions and the relevance and clarity of the

instructions have also been validated (CVR⩾0.37). At this stage, a tool to assess four dimensions (i.e., cognitive, affective, physical, and social) of PL was designed. The first version of the ELIP was constituted with 88 items and three motor tests (Figure 2).

[Figure 2 near here].

Study 2

Cognitive interview procedure. Item readability and comprehension are essential to accuracy in reporting and therefore critical to ensure valid and reliable responses as communication failures are commonplace in questionnaires (Beatty & Willis, 2007). Cognitive interview is a powerful method to understand the thought process used by the students when answering items, and allow the ability to avoid ambiguity, misunderstandings, and identify unfamiliar wording. It helps to ensure that the ELIP's items are clearly understood by the target participants (Beatty & Willis, 2007) and to verify content validity of each question through the emerging adults' perspective. In the present study, cognitive interviews were conducted in the French language and context.

Participants. The cognitive interview procedure was conducted with 32 participants (Table 5). They were enrolled and randomly sampled from three specific populations recruited in three high schools and one university in France. This method allows these three populations in consideration with varying academic profiles to be tested: (1) first-year sports science students; (2) students enrolled in traditional program; (3) and students enrolled in vocational program.

[Table 5 near here].

Data collection. The interviews were conducted individually by a single interviewer (GJ), who was familiar with the study. The interviewer attended training with an expert interviewer and had to conduct four mock interviews before interviews with study participants began. To achieve a high level of attention from the participants, the cognitive interview was

carried out for only one questionnaire per student (from 30 to 45 minutes). According to the consent of the respondents and/or parents, the 32 interviews were audio recorded and transcribed verbatim. We followed Willis' (2005) recommendations to design interview procedures with two steps: (1) introduction and warm-up, and (2) think-aloud and probing techniques. In a first step, a warm-up is conducted to prepare the subject to perform a think-aloud interview: "*Try to visualize the place where you live and think about how many windows there are in that place. Can you give me the number of windows please?*". This question prepares subject to the think aloud method and it illustrates that a poorly formulated or overly general item can be problematic. The interviewer insists on the fact that no perfect answers were expected, but rather that the interview is about testing a questionnaire in progress that includes questions that may be difficult to understand or answer. Participants are reassured by the interviewer's announcement: "*I didn't write these questions, so don't worry, I won't be upset if you criticize this questionnaire. All your comments, whatever they are, can help me*". In a second step, the interviews were conducted through think-aloud technique mixed with probing techniques. The think-aloud method asks respondents to verbalize their thoughts and understanding aloud by reading each item and trying to answer the question as they understand it (Supplementary file 4). Reactive verbal probes were used to question participants in response to his behavior (e.g., "*On this question it took you a long time to answer, what happened when you read the item?*"). Pre-planned probes were used to encourage participants to talk aloud about how they understood, processed, and responded to specific items (e.g., "C*an you understand this word?*"). Moreover, retrospective questions were used to identify different elements that are difficult or blurred (e.g., "*Now that you have read the questionnaire which items were most difficult for you or for one of your friends to complete?*"). In case of misunderstanding, the interviewer tried to get the participant to rephrase the item with his or her own words. Similarly, the probes were also directed to

ensure that the questions asked what they were supposed to measure (e.g., "*Do you think some students will respond '1' here? What would be the difference between you and he/she?*"). This procedure was originally designed in three stages but free to continue until reaching theoretical saturation (i.e., sustaining the process until no new findings emerge). Researchers decide by consensus when this saturation point has been reached (Padilla & Benítez, 2014). To ensure theoretical saturation, the final sets of items were administrated in three other classes of vocational high school (n=68).

Data analysis. The analysis of the cognitive interview data was conducted jointly between two members of the research team (GJ; DT). For each interview question, the two researchers coded impressions to determine whether the participant understood or not the item (Table 6). The analysis of the cognitive interviews data conducted by the two researchers was twofold: (1) the analysis of each misunderstood item and (2) the analysis of each item understood but justified by an unexpected explanation. Based on the problem encoded, the audio recording, and notes, the two experts worked together until they reached a consensus on the decision to be taken (i.e., discard, modify, or keep the item and reformulation). The researchers sought to determine whether an alternative formulation could improve comprehension and what the best possible wording would be (Supplementary file 5). In the case of non-consensus, a third member (PF) contributed to the consensus to identify the best rewording to test in the next step.

Results

Each interview (n=32) was successfully completed without interruption. At the end of each stage, the researchers met to adjust the questionnaire according to the interviews (Supplementary file 5). The cognitive interview data allowed precise questionnaires according to: (1) rephrasing of items not understood or misunderstood based on the participants'

answers (Table 6); and (2) modifying items that did not assess what they were supposed to (Table 6). At the end of this process, no item presented any misunderstanding and thus ensured the content validity of the three questionnaires (i.e., affective, cognitive, and social including the 8 aquatic items). Significant modifications concerning the form of the questionnaire were made (i.e., item headers, highlighted).

At the end of the cognitive interview process, we noted no comments about understanding in the three vocational test classes, and no student refused to answer. The time required to complete the questionnaire was between 10 and 20 minutes for each document. At the end of this phase, we obtained an initial pool of 88 items for testing affective, cognitive, and social dimensions of the ELIP. These items were completed by three motor tests (Figure 1). Each test is presented in Appendix (translated into English for the reader's understanding). **[Table 6 near here].**

Discussion

The aim of this study was to inform how to assess PL for emerging adults (16-21 years old) by developing the foundation and assessing first validation levels of the PL Assessment Tool (ELIP). To overcome this epistemological and methodological challenge, a double process was employed: a Delphi method and a cognitive interview process. At the end of this process, we obtained a comprehensive physical literacy assessment (ELIP) tool useful to measure PL development for emerging adults. The main result showed a global view about PL, according to the 30 experts that participated. For this group, four dimensions were retained: physical, social, affective, and cognitive.

An original tool designed according to four dimensions

To our knowledge, ELIP is the first tool to simultaneously evaluate these four dimensions for emerging adults in French-speaking while, at the same time, considering the philosophical pillars of PL.

Despite the social dimension being already included in robust definitions of PL (Keegan et al., 2019; Martins et al., 2020), it is one of the first PL assessment tools which considers this dimension thoroughly. With the PL-C Quest, PFL and PPLA-Q, ELIP is the first test to focus a full dimension on this issue and specifically within the PA context. (Barnett et al., 2020; Lodewyk, 2019; Mota et al., 2021). Other tests referring to the social dimension (PFL, PPLI) do not orient this dimension in the specific field of PA. To date, ELIP is the only one to include the social dimension as a core component and includes both questionnaires and physical tests. The PL-C Quest and PPLA-Q were designed to map to the Australian Framework (Keegan et al., 2019) and were not available when this investigation began. This focus of the social dimension deviates from Whitehead's well-known definition, which initially did not include it. This selection is probably a consequence of new approaches to PL, included the Australian and European approaches (Keegan et al., 2019, Physical Literacy for Life, 2021) and the specificity of the target population. For emerging adults, the social dimension seems to be a key dimension of the PA commitment process (Lu & Steele, 2014; Van Der Horst et al., 2007). Nevertheless, inclusion of the social domain was first published by Dudley (2015) and early references to social domains were already present in the additional attributes for an optimal PL profile (Whitehead, 2018) (including the ability to work independently and with others in both cooperative and competitive situations). The social dimension of PL needs to be included in future assessment tools especially for this population, for whom social relationships appear to be essential for the quality of life (Edwards et al., 2002).

In view of the affective dimension, ELIP seems to be in line with the previous proposals of 'psychological dimension' (e.g., CAPL). The set of PL tests mentioned above assess this dimension. However, in the ELIP, the affective dimension is distinguished by considering the affective dimension in three different ways: affect towards PA (i.e., pleasure, well-being); affect in PA context (i.e., confidence, self-esteem); management of affect in PA context (i.e., management of emotions). ELIP is one of the recent tools (Barnett et al., 2021) which focus on the management of emotions in PA context. Once again, the expert consensus emphasizes a broader understanding of PL than the four pillars of the Whitehead's definition.

The ELIP cognitive dimension focuses on individual knowledge and cognitive resources and does not deviate too widely from the tools already designed. The originality of the results lies in coupling between Likert Scales assessment and a factual knowledge test. Here, the cognitive dimension goes beyond the fact of 'knowing about PA' but is also concerned with its application and personal awareness through a self-assessment. The importance of this dimension already advocated (Cairney et al., 2019b; Cale & Harris, 2018) is essential to consider for emerging adolescents insofar as it is poorly or not at all supported at school, at least in France (Gandrieau et al., 2021).

Finally, the results of the Delphi method are also original concerning different points of the physical dimension. The ELIP is the first test to broaden the physical dimension by including motor creativity, in addition to cycling, aquatic, motor competences. It's undoubtedly a strong added value of this test, which has yet to be validated by further studies.

Also, the ELIP is innovative in that the fitness dimension is not explicitly present (strength, endurance, etc.). Indeed, the fitness dimension was rejected by the experts who emphasized that is particularly subject to normative comparison and does not represent a fundamental asset for sustainable physical activity education (Rowland, 1995). This perspective clearly differentiates the ELIP from the CAPL and CAEPL which emphasizes the importance

of the fitness dimension. This choice is in line with some criticisms of pre-existing tools (Robinson & Randall, 2017), but opposes some defended views of PL (Tremblay & Longmuir, 2017). The expert panel focused their expectations on aspects of body control rather than physiological resources. It can be argued by its link with the generic term literacy which can be understood as the ability to interact or engage effectively with the environment in which we are situated with our own capabilities. Fitness, from a performance perspective, is not within the scope of most definitions of PL (Shearer et al., 2018). From this point of view ELIP could be more deeply rooted in the philosophy of the concept than other tests including fitness test. In terms of the pedagogical aspect, this choice is significant. In fact, improving fitness can occur with minimal attention to motor competences (Barnett et al., 2021) and ELIP invites PL professionals to lower focus on fitness as a major goal but rather as one of many elements that would result from regular participation in the PA. On this point, the skills needed to engage in resistance or fitness training could be questioned in further considerations. There are already tools to explore this important aspect for emerging adults (Lubans et al., 2014) but experts choose to stand at a more general level for the ELIP. It seems to imply that PL would be placed at a more global level.

The Delphi method also introduced a different way of considering the motor tests of an PL assessment tool. Indeed, the *Star Challenge* test diverges from traditional Fundamental Motor Skills (FMS) (Gallahue et al., 2012) by mobilizing broader movement competences in more complex and open environment (e.g., control body movements in an environment that must be decoded). The participant must perform movement patterns (i.e., jumping, crawling, running, catching, ...) and engage his/her movement capacities (i.e., stability, speed, …) jointly (Durden-Myers et al., 2018) in environments which challenge his/her interpretation. Although the anchoring of FMS in the theory of constraint-led approach has been advocated (Barnett et al., 2016), the selected competences deviate from an isolated evaluation of FMS and emphasize

the importance of "*action intelligence* […] *embedded in perception, experience, memory, anticipation and decision making*" (Whitehead, 2001, p. 131). This specificity gives interesting anchorage with the existentialist perspective. To limit bias (i.e., anxiety, haste) and to best fit the concept of PL, the consideration of time as an evaluation criterion will have to be discussed in more detail according to the motor tasks performed during the test.

The novelty in the physical dimension is also found in the inclusion of cycling and aquatic competences already highlighted by the Australian approach (Barnett et al., 2020; Keegan et al., 2019), and the new perspectives to foundational movement skills (Hulteen et al., 2018). The specific references to the aquatic and cycling world seems relevant in a French holistic PL assessment tool. In fact, aquatic activities are culturally anchored in the practices of French population, representing the 3rd most popular PA category (Croutte & Müller, 2018). Cycling also represents a huge opportunity for PA in France, with 18% of the population involved (Croutte & Müller, 2018). Moreover, both activities represent an opportunity for lifelong commitment in physical activity, as they can be practiced in France across all age and in people with specific health conditions (Leger et al., 2019; Potdevin et al., 2013). Overall, these activities are also widely practiced in Europe and worldwide (Hulteen et al., 2017) and it suggest extending this reflection to other contexts than French context. These two specificities, which have not been considered previously in this way in a PL test, connect PL with numerous beneficial PA opportunities for physical and mental health (Cox et al., 2010; Larouche et al., 2011). Nevertheless, experts seem to stand by a pragmatic position to make the assessment tool easily applicable. They made strong choices regarding water and bike competences. Concerning aquatic competences, the experts chose to focus on perceived rather than objective ones. Therefore, attention must be paid to the gap between these two types of competences to avoid a poor-quality aquatic education. Concerning bike competences, while other wheeled transportations are available (e.g., scooters, skateboards, etc.) the experts chose to focus on

bikes. Other active transportation competences should not be excluded from the physical domain.

Finally, the uniqueness of the physical dimension is also the selection of motor creativity, which has not yet been explored in the PL assessment. Integrating the creativity dimension in a PL tool seems to correspond to the phenomenological perspective in which no precise model is expected but considers the uniqueness of the relationship between the individual and his/her environment that is missing from the existing PL's assessment tools, as well as ecological-dynamics framework perspectives (Rudd et al., 2020). Likewise, Davids et al. (2016) emphasized that creativity is an important component of adaptability. On this point, the expert consensus has stepped back from the well-known Whitehead's definition and thus PL goes beyond confidence, motivation, physical competence, and knowledge and is based on a more comprehensive understanding of the PL concept: "A *disposition to use experience, understanding and abilities to interact effectively with the word*" and enables the emergence of as many answers as possible to a given problem (Whitehead, 2004, p. 4).

These results of the Delphi method suggest that ELIP is based on a broader definition than the Whiteheadian definition without losing its two main axes – the holistic aspect and its orientation towards the promotion of PA for life. The results converge with the Physical Literacy For Life definition (2021), which emphasizes physical, emotional, social, and cognitive skills and attributes. If the distinctions between the definitions can be discussed, it's necessary to place the results of this study in a definition that emphasizes the need for cognitive, social, affective, and physical resources for a sustainable physical lifestyle. As such, we propose that the ELIP could expand and operationalize this definition to identify PL profiles in France and Europe.

A tool designed according to the three philosophical pillars of PL

ELIP is a monitoring and evaluation tool consistent with Whitehead's philosophical perspective anchors. In fact, the existentialist perspective was respected by the wide nature of the PA definition and assessment environments (vs. sports-oriented vision and closed-motor pattern). Items were modified with the dual objectives of "*Physical activity should not be limited to sports (broad movement culture)*" and "*Adolescents should be able to easily project himself/herself into specific experiences that are unique to him/her*". Moreover, the combination of questionnaires with complex objective tests allows to capture an interesting range of information for understanding the unique relationship between the individual and the environment. The complex nature of motor tests is particularly interesting in this respect of existentialism perspective. Now, there is new challenge in considering the singularity of the individual/environment relationship, particularly in the construction of scoring assessment. The complex system approaches (Preiser, 2019) could go beyond the limits announced by the "idealists" (Edwards et al., 2018), opposed to PL assessment, by considering the diversity and uniqueness of PL patterns.

The anchoring in monist perspective is underlined with the holistic consideration of the whole being constituted by four dimensions (i.e., affective, cognitive, physical, and social). ELIP goes beyond the Whiteheadian definition by considering a broad multidimensional range of human components. To further embrace the challenge posed by the monistic perspective, more complex scoring method needs to be explored to consider the interdependence and equal importance of each dimension.

Finally, the challenge of the phenomenological perspective must now be considered in the use of ELIP for emerging adult to be properly addressed. Nevertheless, ELIP has the necessary structure for an accurate ipsative assessment to inform the individuals' own PL journey. The multidimensional nature of ELIP will provide an accurate view of PL development. The tool will need to be used in this sense which some may name as a 'charting'

process (Whitehead, 2019) while optimizing its pragmatic form. Particular attention should also be paid to developing ELIP as an inclusive assessment tool, especially for participants with disabilities.

An empirical and practical tool

ELIP could provide a significant contribution both to research on PL and the practices that underpins it. It will allow measurement of PL in an original way during a critical transition of life (Arnett, 2000) by being more comprehensive than Whitehead's well-known definition (e.g., motor creativity, cycling competences, and management of emotions) but deeply designed in the initial philosophical perspectives. On one hand, ELIP might help teachers' guide and support students in their PL journeys by analyzing student profiles to identify the most vulnerable resources that need to be developed to access more PA opportunities throughout life. Then, ELIP could be a key tool in PE curriculum, particularly on the last year of high school which represents the last year of compulsory PE for the French emerging adults. This tool could encourage the collaboration of PA and health stakeholders around the concept. On the other hand, it will be useful to study the importance of PL during the transition from adolescence to adulthood. The study of this life transition is essential to complete the knowledge of PL value throughout the life course and to understanding the PL journey (Longmuir & Tremblay, 2016).

Limitations and perspectives

A Delphi method with good engagement from PL experts, and the cognitive interviews with students from the targeted age-groups were useful to overcome the difficulties of creating a new educational assessment tool, and specifically to verify face and content validity, as well as assess feasibility. The results of this iterative method must now be tested by a proof of concept to confront the theoretical product to the real complexity of the concept.

It would be interesting to organize a focus group based on the first quantitative results to clarify and confirm the tool being designed. The major challenge will be to finalize the tool with a scoring that best represents the PL concept. Likewise, the design of the tool should be finalized by checking the four methodological pillars: the feasibility, sensitivity, reliability, and validity.

Initial pre-tests will allow us to confirm the psychometrics parameters. The construct validity will be checked by following the COSMIN recommendations by assessing internal and external consistency. Then, criteria validity can be tested by measuring the association between PL and PA (Cairney et al., 2019b). Finally, the stability of the measure can be verified. By continuing to expand the validation of this tool to older and younger populations, it will allow the implementation of charting progress (Green et al., 2018) to assess PL personal journey.

Conclusions

This study allowed the design of a new PL assessment tool conceptually validated by a significant number of international experts and an interview cognitive process. Without questioning the strength of validation of other widely used tools, ELIP complements the literature by providing an original and unique perspective to the field of PL assessment. ELIP is the first tool to investigate PL among emerging adults by combining subjective and motor tests. Its four interdependent dimensions – physical, social, cognitive, and affective – go beyond Whitehead's well-known definition and emphasize a wide range of holistic resources essential for an optimal PL profile. The design methodology emphasized the philosophical anchoring of the tool, but the validation process and progress on feasibility still need to be further measured for a real implementation plan for educators and teachers. The aim of this work is to contribute to the improvement of the measurement tools of the PL and,

590 subsequently, to participate in the promotion and monitoring of the PL among emerging
591 adults.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testin*. American Educational Research Association.

Arnett, J. J. (2000). Emerging adulthood. A theory of development from the late teens through the twenties. *The American Psychologist*, *55*(5), 469–480.

Barnett, L. M., Mazzoli, E., Hawkins, M., Lander, N., Lubans, D. R., Caldwell, S., Comis, P., Keegan, R. J., Cairney, J., Dudley, D., Stewart, R. L., Long, G., Schranz, N., Brown, T. D., & Salmon, J. (2020). Development of a self-report scale to assess children's perceived physical literacy. *Physical Education and Sport Pedagogy*, 1–26. https://doi.org/10.1080/17408989.2020.1849596

Barnett, L. M., Ridgers, N. D., & Salmon, J. (2015). Associations between young children's perceived and actual ball skill competence and physical activity. *Journal of Science and Medicine in Sport*, *18*(2), 167–171. https://doi.org/10.1016/j.jsams.2014.03.001

Barnett, L. M., Stodden, D., Cohen, K. E., Smith, J. J., Lubans, D. R., Lenoir, M., Iivonen, S., Miller, A. D., Laukkanen, A., Dudley, D., Lander, N. J., Brown, H., & Morgan, P. J. (2016). Fundamental Movement Skills: An Important Focus. *Journal of Teaching in Physical Education*, *35*(3), 219–225. https://doi.org/10.1123/jtpe.2014-0209

Barnett, L. M., Webster, E. K., Hulteen, R. M., De Meester, A., Valentini, N. C., Lenoir, M., Pesce, C., Getchell, N., Lopes, V. P., Robinson, L. E., Brian, A., & Rodrigues, L. P. (2021). Through the Looking Glass: A Systematic Review of Longitudinal Evidence, Providing New Insight for Motor Competence and Health. *Sports Medicine*. https://doi.org/10.1007/s40279-021-01516-8

Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, *71*(2), 287–311. https://doi.org/10.1093/poq/nfm006

Berndt, T. J., & Savin-Williams, R. C. (1993). Peer relations and friendships. In *Handbook of clinical research and practice with adolescents.* (pp. 203–219). John Wiley & Sons.

Brasseur, S., Grégoire, J., Bourdu, R., & Mikolajczak, M. (2013). The Profile of Emotional Competence (PEC): Development and Validation of a Self-Reported Measure that Fits Dimensions of Emotional Competence Theory. *PLoS ONE*, *8*(5), e62635. https://doi.org/10.1371/journal.pone.0062635

Cairney, J., Clark, H., Dudley, D., & Kriellaars, D. (2019a). Physical Literacy in Children and Youth—A Construct Validation Study. *Journal of Teaching in Physical Education*, *38*(2), 84–90. https://doi.org/10.1123/jtpe.2018-0270

Cairney, J., Clark, H. J., James, M. E., Mitchell, D., Dudley, D. A., & Kriellaars, D. (2018). The Preschool Physical Literacy Assessment Tool: Testing a New Physical Literacy Tool for the Early Years. *Frontiers in Pediatrics*, *6*, 138. https://doi.org/10.3389/fped.2018.00138

Cairney, J., Dudley, D., Kwan, M., Bulten, R., & Kriellaars, D. (2019b). Physical Literacy, Physical Activity and Health: Toward an Evidence-Informed Conceptual Model. *Sports Medicine*, *49*(3), 371–383. https://doi.org/10.1007/s40279-019-01063-3

Caldwell, H. A., Di Cristofaro, N. A., Cairney, J., Bray, S. R., & Timmons, B. W. (2021). Measurement properties of the Physical Literacy Assessment for Youth (PLAY) Tools. *Applied Physiology, Nutrition, and Metabolism*, *46*(6), 571–578. https://doi.org/10.1139/apnm-2020-0648

Cale, L., & Harris, J. (2018). The Role of Knowledge and Understanding in Fostering Physical Literacy. *Journal of Teaching in Physical Education*, *37*(3), 280–287. https://doi.org/10.1123/jtpe.2018-0134

Chen, A. (2020). A clash of fundamental assumptions: Can/should we measure physical literacy? *Journal of Sport and Health Science*, *9*(2), 149–151. https://doi.org/10.1016/j.jshs.2019.11.002

Chen, S.-T., Tang, Y., Chen, P.-J., & Liu, Y. (2020). The Development of Chinese Assessment and Evaluation of Physical Literacy (CAEPL): A Study Using Delphi Method. *International Journal of Environmental Research and Public Health*, *17*(8), 2720. https://doi.org/10.3390/ijerph17082720

Cox, K. L., Burke, V., Beilin, L. J., & Puddey, I. B. (2010). A comparison of the effects of swimming and walking on body weight, fat distribution, lipids, glucose, and insulin in older women—The Sedentary Women Exercise Adherence Trial 2. *Metabolism*, *59*(11), 1562–1573. https://doi.org/10.1016/j.metabol.2010.02.001

Croutte, P., & Müller, J. (2018). *Baromètre national des pratiques sportives 2018* [Baromètre réalisé par le CREDOC sous la direction de Hoibian S. pour l'INJEP et le ministère des sports]. INJEP Notes & rapports/Rapport d'étude.

Davids, K., Araújo, D., & Brymer, E. (2016). Designing Affordances for Health-Enhancing Physical Activity and Exercise in Sedentary Individuals. *Sports Medicine*, *46*(7), 933–938. https://doi.org/10.1007/s40279-016-0511-3

Dietrich, H., & Ehrlenspiel, F. (2010). Cognitive Interviewing: A Qualitative Tool for Improving Questionnaires in Sport Science. *Measurement in Physical Education and Exercise Science*, *14*, 51–60. https://doi.org/10.1080/10913670903455025

Dudley, D. (2018). Physical Literacy: When the Sum of the Parts Is Greater than the Whole. *Journal of Physical Education, Recreation & Dance*, *89*(3), 7–8. https://doi.org/10.1080/07303084.2018.1418998

Dudley, D. A. (2015). A Conceptual Model of Observed Physical Literacy. *The Physical Educator*. https://doi.org/10.18666/TPE-2015-V72-I5-6020

Durden-Myers, E. J., Green, N. R., & Whitehead, M. E. (2018). Implications for Promoting Physical Literacy. *Journal of Teaching in Physical Education*, *37*(3), 262–271. SPORTDiscus with Full Text.

Edwards, L. C., Bryant, A. S., Keegan, R. J., Morgan, K., Cooper, S.-M., & Jones, A. M. (2018). 'Measuring' Physical Literacy and Related Constructs: A Systematic Review of Empirical Findings. *Sports Medicine*, *48*(3), 659–682. https://doi.org/10.1007/s40279-017-0817-9

Edwards, L. C., Bryant, A. S., Keegan, R. J., Morgan, K., & Jones, A. M. (2017). Definitions, Foundations and Associations of Physical Literacy: A Systematic Review. *Sports Medicine*, *47*(1), 113–126. https://doi.org/10.1007/s40279-016-0560-7

673 Edwards, T. C., Huebner, C. E., Connell, F. A., & Patrick, D. L. (2002). Adolescent quality of life,
674 Part I: Conceptual and measurement model. *Journal of Adolescence*, *25*(3), 275–286.
675 https://doi.org/10.1006/jado.2002.0470
676 Fox, K. R., & Corbin, C. B. (1989). The Physical Self-Perception Profile: Devlopment and
677 Preliminary Validation. *Journal of Sport and Exercise Psychology*, *11*(4), 408–430.
678 https://doi.org/10.1123/jsep.11.4.408
679 Gallahue, D., Ozmun, J., & Goodway, J. (2012). *Understanding Motor Development: Infants,*
680 *Children, Adolescents, Adults. Seventh edn.* (7th ed.).
681 Gandrieau, J., Schnitzler, C., Derigny, T., Lléna, C., Mouton, A., & Potdevin, F. (2021). How
682 Knowledge About Physical Activity Is Impacted By School Institution, Grade Level, and
683 Gender Throughout High School Years in France? *Journal of Teaching in Physical Education*,
684 1–9. https://doi.org/10.1123/jtpe.2021-0092
685 Green, N. R., Roberts, W. M., Sheehan, D., & Keegan, R. J. (2018). Charting Physical Literacy
686 Journeys Within Physical Education Settings. *Journal of Teaching in Physical Education*,
687 *37*(3), 272–279. https://doi.org/10.1123/jtpe.2018-0129
688 Gunnell, K. E., Schellenberg, B. J. I., Wilson, P. M., Crocker, P. R. E., Mack, D. E., & Zumbo, B. D.
689 (2014). A Review of Validity Evidence Presented in the Journal of Sport and Exercise
690 Psychology (2002–2012): Misconceptions and Recommendations for Validation Research. In
691 B. D. Zumbo & E. K. H. Chan (Eds.), *Validity and Validation in Social, Behavioral, and*
692 *Health Sciences* (Vol. 54, pp. 137–156). Springer International Publishing.
693 https://doi.org/10.1007/978-3-319-07794-9_8
694 Hanson, C. L., Oliver, E. J., Dodd-Reynolds, C. J., Pearsons, A., & Kelly, P. (2020). A modified
695 Delphi study to gain consensus for a taxonomy to report and classify physical activity referral
696 schemes (PARS). *International Journal of Behavioral Nutrition and Physical Activity*, *17*(1),
697 158. https://doi.org/10.1186/s12966-020-01050-2
698 Harter, S. (2017). *Self-Perception Profile for Adolescents* [Data set]. American Psychological
699 Association. https://doi.org/10.1037/t05703-000

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Heino, M. T. J., Knittle, K., Noone, C., Hasselman, F., & Hankonen, N. (2021). Studying Behaviour Change Mechanisms under Complexity. *Behavioral Sciences*, *11*(5), 77. https://doi.org/10.3390/bs11050077

Hulteen, R. M., Morgan, P. J., Barnett, L. M., Stodden, D. F., & Lubans, D. R. (2018). Development of Foundational Movement Skills: A Conceptual Model for Physical Activity Across the Lifespan. *Sports Medicine*, *48*(7), 1533–1540. https://doi.org/10.1007/s40279-018-0892-6

Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P. C., Colyvas, K., & Lubans, D. R. (2017). Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. *Preventive Medicine*, *95*, 14–25. https://doi.org/10.1016/j.ypmed.2016.11.027

IPLA. (2017). *IPLA definition. Retrieved from*. https://www.physical-literacy.org.uk/

Keegan, R. J., Barnett, L. M., Dudley, D. A., Telford, R. D., Lubans, D. R., Bryant, A. S., Roberts, W. M., Morgan, P. J., Schranz, N. K., Weissensteiner, J. R., Vella, S. A., Salmon, J., Ziviani, J., Okely, A. D., Wainwright, N., & Evans, J. R. (2019). Defining Physical Literacy for Application in Australia: A Modified Delphi Method. *Journal of Teaching in Physical Education*, *38*(2), 105–118. https://doi.org/10.1123/jtpe.2018-0264

Kwan, M. Y. W., Graham, J. D., Bedard, C., Bremer, E., Healey, C., & Cairney, J. (2019). Examining the Effectiveness of a Pilot Physical Literacy–Based Intervention Targeting First-Year University Students: The PLUS Program. *SAGE Open*, *9*(2), 215824401985024. https://doi.org/10.1177/2158244019850248

Larouche, R., Lloyd, M., Knight, E., & Tremblay, M. S. (2011). Relationship Between Active School Transport and Body Mass Index in Grades—4-to-6 Children. *Pediatric Exercise Science*, *23*(3), 322–330. SPORTDiscus with Full Text.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leger, S. J., Dean, J. L., Edge, S., & Casello, J. M. (2019). “If I had a regular bicycle, I wouldn’t be out riding anymore”: Perspectives on the potential of e-bikes to support active living and independent mobility among older adults in Waterloo, Canada. *Transportation Research Part A: Policy and Practice*, *123*, 240–254. https://doi.org/10.1016/j.tra.2018.10.009

Li, M. H., Sum, R. K. W., Sit, C. H. P., Liu, Y., & Li, R. (2021). Perceived and actual physical literacy and physical activity: A test of reverse pathway among Hong Kong children. *Journal of Exercise Science & Fitness*, *19*(3), 171–177. https://doi.org/10.1016/j.jesf.2021.03.001

Liu, Y., & Chen, S. (2020). Physical literacy in children and adolescents: Definitions, assessments, and interventions. *European Physical Education Review*, 1356336X2092550. https://doi.org/10.1177/1356336X20925502

Lodewyk, K. R. (2019). Early Validation Evidence of the Canadian Practitioner-Based Assessment of Physical Literacy in Secondary Physical Education. *Physical Educator*, *76*(3), 634–660. SPORTDiscus with Full Text.

Longmuir, P. E., Boyer, C., Lloyd, M., Yang, Y., Boiarskaia, E., Zhu, W., & Tremblay, M. S. (2015). The Canadian Assessment of Physical Literacy: Methods for children in grades 4 to 6 (8 to 12 years). *BMC Public Health*, *15*(1), 767. https://doi.org/10.1186/s12889-015-2106-6

Longmuir, P. E., & Tremblay, M. S. (2016). Top 10 Research Questions Related to Physical Literacy. *Research Quarterly for Exercise and Sport*, *87*(1), 28–35. https://doi.org/10.1080/02701367.2016.1124671

Lu, C., & Steele, K. (2014). Characteristics and conceptual framework of the Easy-Play Model. *European Physical Education Review*, *20*(2), 229–243. https://doi.org/10.1177/1356336X14524861

Lubans, D. R., Smith, J. J., Harries, S. K., Barnett, L. M., & Faigenbaum, A. D. (2014). Development, test-retest reliability, and construct validity of the resistance training skills battery. *Journal of Strength and Conditioning Research*, *28*(5), 1373–1380. https://doi.org/10.1519/JSC.0b013e31829b5527

Lundvall, S. (2015). Physical literacy in the field of physical education – A challenge and a possibility. *Journal of Sport and Health Science*, *4*(2), 113–118. https://doi.org/10.1016/j.jshs.2015.02.001

Markland, D., & Tobin, V. (2004). A Modification to the Behavioural Regulation in Exercise Questionnaire to Include an Assessment of Amotivation. *Journal of Sport and Exercise Psychology*, *26*(2), 191–196. https://doi.org/10.1123/jsep.26.2.191

Martins, J., Onofre, M., Mota, J., Murphy, C., Repond, R.-M., Vost, H., Cremosini, B., Svrdlim, A., Markovic, M., & Dudley, D. (2020). International approaches to the definition, philosophical tenets, and core elements of physical literacy: A scoping review. *PROSPECTS*. https://doi.org/10.1007/s11125-020-09466-1

Merleau-Ponty, M. (1968). *Phenomenology of perception (C. Smith, Trans.)*. Routledge.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Moran, K., Stallman, R. K., Kjendlie, P.-L., Dahl, D., Blitvich, J. D., Petrass, L. A., McElroy, G. K., Goya, T., Teramoto, K., Matsui, A., & Shimongata, S. (2012). Can You swim? An Exploration of Measuring Real and Perceived Water Competency. *International Journal of Aquatic Research and Education*, *6*(2). https://doi.org/10.25035/ijare.06.02.04

Mota, J., Martins, J., & Onofre, M. (2021). Portuguese Physical Literacy Assessment Questionnaire (PPLA-Q) for adolescents (15–18 years) from grades 10–12: Development, content validation and pilot testing. *BMC Public Health*, *21*(1), 2183. https://doi.org/10.1186/s12889-021-12230-5

Muller, J. (2018). *Les jeunes aiment le sport...de préférence sans contraintes* (No. 297; p. 3). UCPA/CREDOC.

Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *26.1*, 136–144. https://doi.org/10.7334/psicothema2013.259

Physical Literacy For Life. (2021). https://physical-literacy.isca.org

Pot, N., Whitehead, M. E., & Durden-Myers, E. J. (2018). Physical Literacy From Philosophy to Practice. *Journal of Teaching in Physical Education*, *37*(3), 246–251. https://doi.org/10.1123/jtpe.2018-0133

Potdevin, F., Normani, C., & Pelayo, P. (2013). Examining Self-Training Procedures in Leisure Swimming. *Journal of Sports Science & Medicine*, *12*, 716–723.

Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, *41*(4), 376–382. https://doi.org/10.1046/j.1365-2648.2003.02537.x

Preiser, R. (2019). Identifying general trends and patterns in complex systems research: An overview of theoretical and practical implications. *Systems Research and Behavioral Science*, *36*(5), 706–714. https://doi.org/10.1002/sres.2619

Robinson, D. B., & Randall, L. (2017). Marking Physical Literacy or Missing the Mark on Physical Literacy? A Conceptual Critique of Canada's Physical Literacy Assessment Instruments. *Measurement in Physical Education and Exercise Science*, *21*(1), 40–55. https://doi.org/10.1080/1091367X.2016.1249793

Rowland, T. W. (Ed.). (1995). "The Horse Is Dead; Let's Dismount." *Pediatric Exercise Science*, *7*(2), 117–120. https://doi.org/10.1123/pes.7.2.117

Rudd, J. R., Pesce, C., Strafford, B. W., & Davids, K. (2020). Physical Literacy - A Journey of Individual Enrichment: An Ecological Dynamics Rationale for Enhancing Performance and Physical Activity in All. *Frontiers in Psychology*, *11*, 1904. https://doi.org/10.3389/fpsyg.2020.01904

Sallis, J. F., Pinski, R. B., Grossman, R. M., Patterson, T. L., & Nader, P. R. (1988). The development of self-efficacy scales for healthrelated diet and exercise behaviors. *Health Education Research*, *3*(3), 283–292. https://doi.org/10.1093/her/3.3.283

Shearer, C., Goss, H. R., Edwards, L. C., Keegan, R. J., Knowles, Z. R., Boddy, L. M., Durden-Myers, E. J., & Foweather, L. (2018). How Is Physical Literacy Defined? A Contemporary Update. *Journal of Teaching in Physical Education*, *37*(3), 237–245. https://doi.org/10.1123/jtpe.2018-0136

Sum, R. K. W., Ha, A. S. C., Cheng, C. F., Chung, P. K., Yiu, K. T. C., Kuo, C. C., Yu, C. K., & Wang, F. J. (2016). Construction and Validation of a Perceived Physical Literacy Instrument for Physical Education Teachers. *PLOS ONE*, *11*(5), e0155610. https://doi.org/10.1371/journal.pone.0155610

Tremblay, M. S., & Longmuir, P. E. (2017a). Conceptual Critique of Canada's Physical Literacy Assessment Instruments Also Misses the Mark. *Measurement in Physical Education & Exercise Science*, *21*(3), 174–176.

Tremblay, M. S., & Longmuir, P. E. (2017b). Conceptual Critique of Canada's Physical Literacy Assessment Instruments Also Misses the Mark. *Measurement in Physical Education and Exercise Science*, *21*(3), 174–176. https://doi.org/10.1080/1091367X.2017.1333002

UNESCO. (2005). *Understanding and defining literacy*. https://unesdoc.unesco.org/ark:/48223/pf0000146186

Van Der Horst, K., Paw, M. J. C. A., Twisk, J. W. R., & Van Mechelen, W. (2007). A brief review on correlates of physical activity and sedentariness in youth. *Medicine and Science in Sports and Exercise*, *39*(8), 1241–1250. https://doi.org/10.1249/mss.0b013e318059bf35

Wang, C. K. J., Liu, W. C., Biddle, S. J. H., & Spray, C. M. (2005). Cross-cultural validation of the Conceptions of the Nature of Athletic Ability Questionnaire Version 2. *Personality and Individual Differences*, *38*(6), 1245–1256. https://doi.org/10.1016/j.paid.2004.08.007

Whitehead, M. (2001). The Concept of Physical Literacy. *European Journal of Physical Education*, *6*(2), 127–138. https://doi.org/10.1080/1740898010060205

Whitehead, M. (2007). Physical Literacy: Philosophical Considerations in Relation to Developing a Sense of Self, Universality and Propositional Knowledge. *Sport, Ethics and Philosophy*, *1*(3), 281–298. https://doi.org/10.1080/17511320701676916

Whitehead, M. (2010). *Physical Literacy: Throughout the Lifecourse*. Routledge. https://doi.org/10.4324/9780203881903

Whitehead, M. (2019). *Physical Literacy across the World*. Routledge. https://doi.org/10.4324/9780203702697

Whitehead, M. E. (2004). PHYSICAL LITERACY - A DEBATE. (Abstract). In *In Klisouras, V. (ed.), 2004 PRE-OLYMPIC CONGRESS: Proceedings: VOLUME I: LECTURES-ORALS: SPORT SCIENCE THROUGH THE AGES, 6-11 AUGUST 2004, THESSALONIKI/HELLAS, ARISTOTLE UNIVERSITY CAMPUS, Thessaloniki, Greece, Aristotle University of Thessaloniki, Department of Physical Education & Sport Science, [2004], p.117-118.*

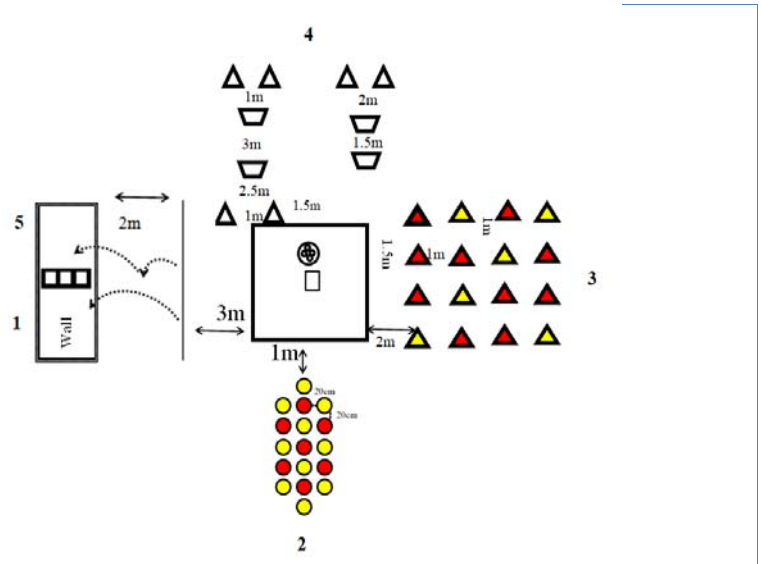
Willis, G. B. (2005). *Cognitive Interviewing*. SAGE Publications Inc. https://us.sagepub.com/en-us/nam/cognitive-interviewing/book225856

Young, L., O'Connor, J., & Alfrey, L. (2020). Physical literacy: A concept analysis. *Sport, Education and Society*, *25*(8), 946–959. https://doi.org/10.1080/13573322.2019.1677586

Zhu, W. (2012). *Measurement Practice in Sport and Exercise Psychology: A Historical, Comparative, and Psychometric View*. https://doi.org/10.5040/9781492596332.ch-002

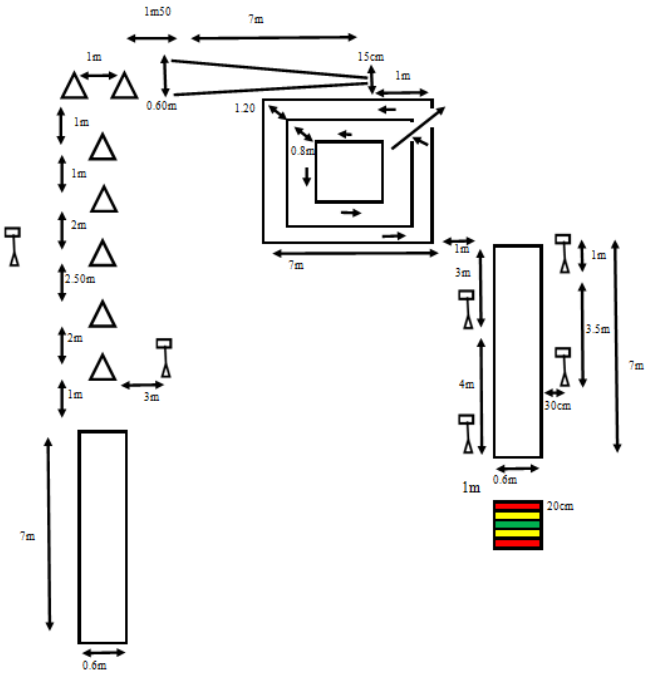
853

854 Appendix 1. The *Star Challenge* (SC)



N°	Name	Competencies assessed	Instructions
1	Direct Targets	Controlling an object - direct throwing and catching	Hit each target with the balls with direct throws and receiving each one. A ball can only be used once. You must validate the first target (top) before validating the 2nd (middle) and then the 3rd (bottom).
2	The water lily pond	Moving while maintaining a dynamic balance	Move as fast as possible on bell-foot only on the yellow markings without touching the red markings or putting the 2nd feet on the ground. Come back with other feet.
3	The labyrinth	Controlling one's body in an environment to be interpret	Touch by hand all the yellow cones without touching the others and do not drop them. The cones are in a precarious balance
4	Hurdles	Moving while crossing obstacles	Run as fast as possible by crossing obstacles without touching them.
5	Bounce Targets	Controlling an object – skip-shot and catching	Hit each target with the balls with skip-shot and receiving each one. A ball can only be used once. You must validate the first target (top) before validating the 2nd (middle) and then the 3rd (bottom).
At the beginning, touch the digital tablet to start the timer. After each challenge, come back and touch the tablet to start the intermediate timer and remember the next challenge.			

855 Appendix 2. The *Bike Challenge* (BC)



N°	Name	Competencies assessed	Instructions
1	Start in track	Starting and produce speed while maintaining heading	Get on the bike and start riding while staying in the track.
2	The slalom	Maintaining a winding trajectory without continuous visual control	Cross the slalom around each cone. Before the end of the slalom (the front wheel passes between the last two cones), turn your head to read and announce the two numbers on the boards.
3	The funnel	Maintaining a controlled and precise trajectory	Drive into the funnel without getting out the lines.
4	The squares	Maintaining direction in a turn	Drive around the squares by staying on the track.
5	Touch and stay in track	One handed controlling	Drive straight into the hallway without getting out and touch the 4 targets with hand.
6	Stop	Stopping suddenly with precision and balance	Stop the front wheel precisely in the specified green area on the ground.

856

Appendix 3. The *Creativity Challenge* guidelines (CC)

1- *Challenge 1: At the signal, you have up to one minute to make as many crossings as possible and in different ways each time.*

2- *Challenge 2: It is a stick, but what else could it be? You have up to one minute to show whatever it might be.*

Appendix 4. *Aquatic Competences Test* (ACT)

					
How many lengths of a 25-meter pool can you swim without stopping or touching the floor?	Less than 2 lengths (50m)	Between 2 and 4 lengths (between 50 and 100m)	Between 4 and 8 lengths (between 100 and 200m)	Between 8 and 12 lengths (between 200 and 300m)	More than 12 lengths (more than 300m)
Can you float in the water without flotation equipment? (arm and leg movements are allowed)	No, it is impossible	Yes, less than 2 minutes	Yes, between 2 and 6 minutes	Yes, between 6 and 15 minutes	Yes, at least 15 minutes
Would you be able to float for the same time in open and deep water (sea, lake)	No, it is impossible	With great difficulty	Without much difficulty	Easily	Very easily
Can you dive safely in a pool?	Impossible to dive in headfirst	Yes, by squatting on the border	Yes, standing on the border	Yes, often from a small diving board (1m)	Yes, I am totally comfortable on the small diving board (1m)
Can you swim underwater in a pool?	No, I can't go underwater	No, but I can go underwater a bit	Yes, I can go about 5 to 10 meters under water	Yes, I can go about 10 to 15 meters under water	Yes, I can go 15 meters or more underwater
Can you look for an item underwater, without diving and headfirst?	No, it is impossible	Only if I can stand in the water	Yes, between 1 and 2 meters deep in a pool	Yes, between 2 and 3 meters deep in a pool	Yes, even at a depth of 3 meters (sea, lake, ...)
Can you swim 100 meters (4 lengths) on your back without stopping?	No, it is impossible	With great difficulty	Without much difficulty	Easily	Very easily
When you are in the water, you would say that...	I am still in great difficulty	I am in trouble if I don't have a foothold	I can have difficulties in great depth	I am quite comfortable in deep water	I am extremely comfortable in deep water

866 Appendix 5. *Cognitive Dimension Survey* (CDS) (Likert Scale).
867

You are in a playground to practice a physical activity. Finally, you realize you are too many to play. Do you think you will be able to propose a new organization to continue practicing?
You are doing a physical activity and your motivation is decreasing. Do you think you will be able to think for a solution to modify what you are practicing while having fun?
You are performing a physical activity. You or your friends are having great difficulty. Do you think you will be able to come up with a solution so that everyone can continue to practice while having fun?
During a busy week, do you think you will be able to modify your planned physical activity to match the time available and your tiredness?
Friends suggest you try a new physical activity. Do you think you have the basic knowledge to easily understand the rules?
I know why physical activity is essential for my health
I sincerely believe that it is necessary to have enough physical activity
I know that physical activity and movement are especially important in life
I understand why it is necessary to make physical exercise and movements
I really know why it is necessary to be physically active every day
I believe that physical activity and movement are absolutely not a waste of time in life
No matter what happens, I will continue to believe that physical activity is a particularly important part of life
I know that physical activity is an essential part of wellbeing
Link the proposed activities with the corresponding category: Not beneficial / Beneficial / Greatly beneficial (24 items)
Identify and check off the signs that you can usually identify that you have engaged in vigorous physical activity
True or false? (23 items)
In your opinion, between the ages of 18 and 64, how many minutes of moderate physical activity per week are recommended as a minimum to maintain health?
In your opinion, between the ages of 18 and 64, how many minutes of vigorous physical activity per week are recommended as a minimum to maintain good health?
From ages 5 to 17, how much moderate physical activity per day do you think is recommended as a minimum to maintain health?
In your opinion, how many steps per day is it recommended to take to maintain health?
What is the minimum duration for a physical activity to be considered beneficial to health?

868

Appendix 6. *Social Dimension Survey* (SDS) (Likert Scale).

I usually find it easy to make friends when I am practicing a physical activity
I usually manage to behave well in a group during physical activities
Generally, I can easily organize a physical activity with a group of people
Generally, I can easily resolve conflicts in a group to be able to continue the physical activities
Generally, I can easily play, practice, or train with other people
Generally, I can easily observe and discuss with someone to learn a movement or a physical activity
I generally find it easy to ask for help or support when I need it in a physical activity
Generally, I can easily come to others to help or encourage them in physical activity
I can usually get someone to learn a physical activity by demonstrating and explaining
Generally, others behave nicely with me in physical activities
I think I am well appreciated by other people my age in physical activity
Generally, I am easily accepted by others in physical activities
Most of the time I feel very well integrated by others in physical activities
I try to make everyone feel really good about physical activities
I respect and enforce the rules of a physical activity so that everyone can practice without injustice
I make sure that no one is excluded from physical activities because everyone has the right to practice
Sometimes, I do not think only about myself and I also think about others so that the physical activity goes well
Sometimes in physical activities, I try to understand how others feel so that they enjoy the practice more
I feel supported by my family, teachers, or friends in the physical activities I want to do
I feel that my loved ones encourage me to be physically active
I know where to go to practice physical activity when I want to
I can find physical activities that I like in what is offered to me (at school, in clubs, by my family or friends, ...)

Appendix 7. *Affective Dimension Survey* (ADS) (Likert Scale).

Generally, when I practice a physical activity, I spend an enjoyable moment
Generally, when I practice physical activity, I feel good
Generally, when I practice a physical activity, I get something positive out of it
Generally, when I practice a physical activity, I am satisfied with what I achieve
Generally, no matter what happens (bad weather, unforeseen events, ...), I enjoy physical activities
Personally, I like physical activity
Personally, physical activity really contributes to my happiness
Personally, I enjoy the different physical activities in my day or week
Personally, I often take pleasure in physical activity
Generally, when I practice physical activity, I feel good about myself
Generally, when I practice physical activity, I am happy with who I am
Generally, when I am physically active, I am happy with what I can achieve
Generally, when I am physically active, I am able to think positively
Generally, when I discover a physical activity that I have never tried before, I am confident to participate
Personally, I am generally comfortable with physical activities
Personally, I am often successful in the physical activities I try
Personally, I perform well in many physical activities
Generally, I accept the physical challenges offered to me
Generally, even after a mistake, I still believe that I can succeed in my physical activity
Generally, I am confident in my ability to succeed in the physical activities I do
Generally, I have enough confidence to try physical activities that I am not familiar with
Generally, when I am confronted with a new physical activity, I do not feel very confident
I am confident enough in my abilities to engage in most physical activities in the water (swimming, kayaking, water games, ...)
I have enough confidence in my abilities to engage in most physical activities in the air (climbing, acrobatic jumps, zip lines, ...)
I am confident enough in my abilities to engage in most physical activities on land (indoor or outdoor)
I am confident enough in my abilities to engage in most physical activities in the snow or on ice (skiing, skating, sledding, ...)
Generally, when I am in physical activity and after a mistake, I remain confident enough to continue my practice
Generally, when I practice physical activity, I am confident in what I am doing
When I practice physical activity, I can easily recover my composure after an irritating event
When I practice physical activity, I can't talk calmly with others
When I practice physical activity, I control my emotions, even in difficult moments
Generally, when I'm sad, it's easy for me to get back into a good mood so I can continue my physical activity with pleasure
Generally, when I am angry in my physical activities, I can easily calm down to resume my practice in all serenity
Generally, when I fail to reach my goal in a physical activity, I am even more motivated to continue to succeed
Generally, physical activity is a very good way to put me in a good mood after negative events

875

876 Table 1. Descriptive Characteristics of Experts Completing the Delphi Process

Characteristics	**Description**	***n*=30**
Gender	Male	21
	Female	9
Location	Western Europe	15
	Southern Europe	4
	Northern Europe	1
	Eastern Europe	1
	Oceania	5
	North America	3
	Asia	1
Area of expertise open-ended question with multiple responses	Physical activity	18
	Physical literacy	14
	Psychology	1
	Physical education	10
	Health education	3
	Sport pedagogy	7
	Motor competences / movement sciences	10
	Sociology	1
	Physical Self-perception	1
	Fitness	1
	Child/adolescents	2
	PE teacher / coach development	4
Career length Self-report	Mean	18.92 years
	Range	3-45 years
PL expertise Self-report	Mean	4.63/5
	Standard deviation	0.49

877

878

879

880 Table 2. Results from round 1 and 2 of the Delphi process

Sub-round 1				
Dimension	**Agreement (%)**	**Domains**	**Agreement (%)**	**Retained/discarded**
Affective	100%	Motivation	100%	Retained
		Self Esteem	95.6%	Retained
		Emotional competence	73.1%	Retained
		Belief	65.2%	Re-presented in sub-round 2
		Confidence	86.9%	Retained
		Enjoyment	91.3%	Retained
		Effort	56.5%	Discarded
Cognitive	95.6%	Benefits and risks	82.6%	Retained
		Recommendations	78.2%	Retained
		Training	56.5%	Discarded
		Principles of efficacy	30.4%	Discarded
		Sports culture	34.7%	Discarded
Physical	100%	Objective tests	82.6%	Retained
		Subjective tests	17.4%	Discarded
		Movement competencies	91.3%	Retained
		Perceived Aquatic competencies	73.9%	Retained
		Cardio	65.2%	Re-presented in sub-round 2
		Bicycle competencies	69.5%	Retained
		Motor creativity	69.5%	Retained
		Strength	65.2%	Re-presented in sub-round 2
Social	86.9%	Social acceptance	69.5%	Retained
		Relationship attractiveness	13.0%	Discarded
		Behavioral conduct	43.5%	Discarded
		Intimate friendship	39.1%	Discarded
		Relational skills	78.3%	Retained
		Self-awareness skills	43.5%	Discarded
		Self-management	82.6%	Retained
		Social awareness	73.9%	Retained
		Responsible decision-making	86.9%	Retained
Sub-round 2				
		Physical fitness (strength, endurance)	56.5%	Discarded
		Belief	78.2%	Retained for cognitive dimension

881

882

Table 3. Results from the third round of the Delphi process

	Sub-round 1		
PL domains	**Tests**	**Agreement (%)**	**Retained/discarded**
Motivation	BREQ-2 (1)	79.2%	Retained
Self-esteem	PSPP (2)	70.8%	Retained
Emotional competence	PEC (3)	62.5%	Re-presented for a 2nd sub-round
Belief	CNAAQ-2 (4)	50%	Discarded
Confidence	Bopp & Vadeboncoeur (5)	70.8%	Retained
Confidence	Self-efficacy scales (6)	75%	Retained
Enjoyment	PAES (7)	79.1%	Retained
Knowledge	KPA (8)	70.8%	Retained
Motor competencies	Star Challenge (9)	70.8%	Retained
Aquatic competencies	Can you swim? (10)	62.%	Re-presented for a 2nd sub-round
Bike competencies	Bike Challenge (11)	70.8%	Retained
Motor creativity	Creativity test (12)	62.%	Re-presented for a 2nd sub-round
Social domains	SSIS (13)	66.6%	Retained
Social domains	SPPA (14)	66.6%	Retained
	Sub-round 2		
Emotional competence	PEC	78.5%	Items intra-personal only
Aquatic competencies	Can you swim?	74%	8 items selected
Motor creativity	Creativity test	79.5%	completed by the ICM test (15)

(1) Markland & Tobin, 2004 (2) Fox & Corbin, 1989 (3) Brasseur et al., 2013 (4) Wang et al., 2005 (5) Bopp & Vadeboncoeur, 2019 (6) Sallis et al., 1988 (7) Kendzierski & Decarlo, 1991 (8) Knowledge of Physical Activity (personal proposal) (9) personal proposal (10) Moran et al., 2012 (11) personal proposal (12) personal proposal (13) Gresham & Elliot, 2008 (14) Harter, 2017 (15) Méndez–Martínez & Fernández–Río, 2019

Commenté [JG1]: problem in zotero, I will complete before submission.

889

890 Table 4. Results from round 4 of the Delphi process

Dimension	Filters	Agreement (%)	Retained/ Discarded
Social	The social dimension should focus on physical activity context (not to be generic)	92%	Retained
	Collective work skill, cooperation, and collaboration need to be more present in social items	92%	Retained
	Bronfenbrenner's model can be used to diversify levels (individual, friends, family, society)	72%	Retained
	Physical activity should not be limited to sports (broad movement culture)	96%	Retained
	Social items can be linked directly with another dimension (e.g., social competencies and enjoyment in physical activity)	52%	Discarded
Affective	Affective dimension needs to be causally related to physical activity	92%	Retained
	Physical activity should not be limited to sports (broad movement culture)	96%	Retained
	This culture of movement must be able to represent all the reasons for engaging in physical activity	100%	Retained
	Adolescents should be able to easily project himself/herself into specific experiences that are unique to him/her	76%	Retained
	Bronfenbrenner's model can be used as inspiration to diversify levels	68%	Retained
	Affective dimension can be directly linked with other dimension (e.g., motivation in PA and physical competencies)	52%	Discarded
Cognitive	Cognitive items need to be less focused on factual knowledge and more focused on cognitive facilitators of an active lifestyle	96%	Retained
	PA should not be limited to sports (broad movement culture)	100%	Retained
	Beliefs (initially proposed in the affective dimension) will be included in this dimension and focus on the importance of PA in life	76%	Retained
	Cognitive dimension could be constituted of 3 domains: (1) beliefs/reasoning about PA (2) factual knowledge (3) manipulation of rules	76%	Retained
Physical	Rank 3 domains (bike/water/land) to weight the physical dimension score	56%	Discarded

891

892

893

894 Table 5. Descriptive Characteristics of Cognitive interview participants

Questionnaire	Characteristics	Number
Affective (*n*=11)	Male	3
	Female	8
	Sports sciences student	4
	Traditional school program	4
	Vocational school program	3
	Age (mean)	17.63(±.67)
Cognitive (*n*=10)	Male	3
	Female	7
	Sports sciences student	3
	Traditional school program	4
	Vocational school program	3
	Age (mean)	17.60(±.83)
Social (*n*=11)	Male	4
	Female	7
	Sports sciences student	4
	Traditional school program	4
	Vocational school program	3
	Age (mean)	17.45(±.69)

895

896

Table 6. Examples of processing an observation during the cognitive interview (from a translation from French to English in these examples).

Example 1	**Initial item read by the emerging adult:** *Personally, physical activity contributes to my fulfilment.* **Analysis of reactions by researchers:** Two of the three participants had difficulty interpreting what was meant by the item: "*Now I didn't understand*"; "*What is fulfilment?*". The third respondent answered it easily. **Probes used by researchers:** "*If I wasn't here, how would you understand the sentence?*" "*What word could you use instead?*" **Summary of the answers of the emerging adult:** Participants prefer the term of "happiness". **Consensus for modification of the item:** The two researchers chose to replace the word "*fulfilment*" with the word "*happiness*", underlining it with "*really*". **Formulation of the item at the end of the stage:** *Personally, physical activity really contributes to my happiness.*
Example 2	**Initial item read by the emerging adult:** *Usually, I can easily make friends when I practice physical activities.* **Probes used by researchers:** *"Do you think any students are going to answer 1 or 2 here? and why*?" **Summary of the answers of the emerging adult:** The 3 participants make a direct link with social skills: "*If people are not comfortable with others, they can choose 1*"; "Yes *in relation to shyness with others and all"*; "Y*es it depends on the ability to integrate."* **Consensus for modification of the item:** The researchers validate that the question is consistent with what it was supposed to measure. **Formulation of the item at the end of the stage:** *Usually, I can easily make friends when I practice physical activities*

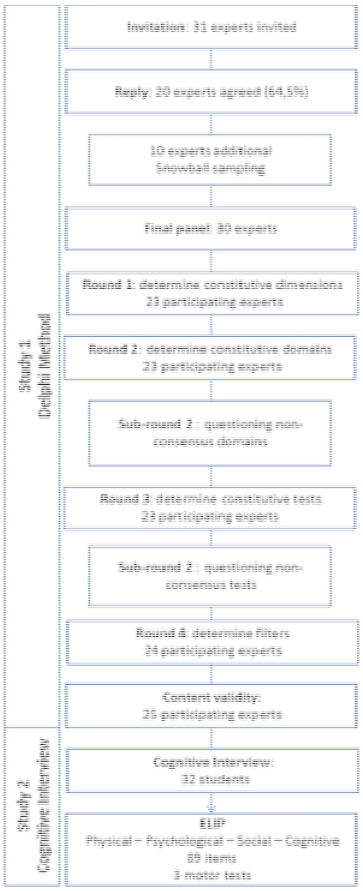


Figure 1. The procedure of ELIP's design

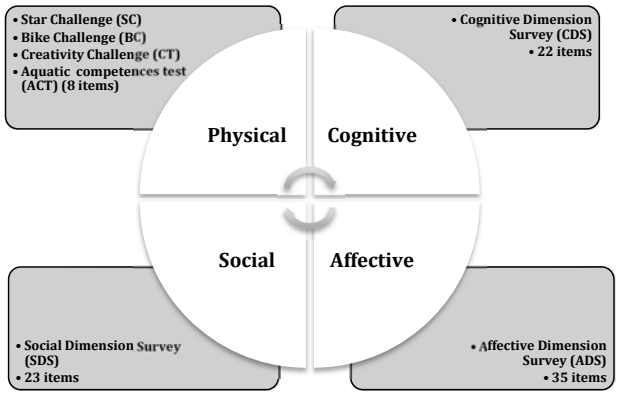


Figure 2. ELIP's Design. *SC (Appendix 1). BC (Appendix 2). CT (Appendix 3). ACT (Appendix 4). CDS (Appendix 5). SDS (Appendix 6). ADS (Appendix 7).*

Response to reviewers

Development of ELIP (Évaluation de La LIttératie Physique) to assess physical literacy for emerging adults: a challenge embraced through Delphi method and cognitive interview process.

Dear Dr. Mark Williams, Editor-in-Chief, Research Quarterly for Exercise and Sport

We would like to thank you and the two reviewers for the careful reading of our work and their relevant advice to improve our manuscript, which has been rewritten.

We really appreciate the positive feedbacks in relation regarding our methodology and the tool designed. We are delighted that you have identified the strengths and potential of this study to contribute to the field.

We also thank the evaluators for their constructive feedbacks which will allow us to improve the quality of this manuscript including the clarity of our philosophical anchoring and the specificity of the tool.

We have answered point by point to the comments. As requested in your letter, we highlighted and used different colours to highlight the location of our changes.

Since we have considered all the comments of the reviewers and in order to maintain the coherence of the manuscript, we have sometimes taken the liberty of modifying very slightly other elements (addition of subheadings, order of sentences). The reviewer's constructive comments improved the manuscript in this way.

We hope this new version corresponds to the attempt of the reviewers and can be considered for publication in RQES.

Sincerely yours,

The authors.

Response to reviewer 1

Reviewer's comments	Response to reviewer	Effective changes
1 I recommend this acceptance of this manuscript. While I would question the physical assessment included in the ELIP, this research advances the understanding of physical literacy and for some addresses a gap. I will make a few comments that the authors may want to consider in finalizing their submission to Research Quarterly for Exercise and Sport.	Thank you very much for your time, consideration, and constructive comments. We have taking them into consideration when revising the manuscript and helped us in improving it.	The consideration of each comment is located at different places in the text with different colors. We have also specified the page and the line in this column.
2 As is discussed, the state of PL testing worldwide is in constant development and dispute. With ELIP and other assessments my questions are what is the purpose of the PL assessment and what is done with the data? You state, "*To forward the development of such tools, the objective of this study was to build the foundations of ELIP: that has been designed to help reduce tensions existing approaches to PL assessment resulting in a low uptake in applied settings*." As I read the manuscript, I wonder who will use the ELIP, in what setting, and for what purposes? You may want to comment on this.	Thank you for this comment, we have clarified this issue in our manuscript. There is a real debate around the evaluation of PL, hence the need to propose a consensual method to reach a consensus. This is particularly important since no tool focuses on this age group. The ELIP was constructed with the objective of being more consistent with the philosophy of PL compared to other measures used in previous studies. The purpose of the ELIP is to propose a new way to assess PL for two purposes. We have detailed these ideas in our paper. - First, our objective is to construct a useful PL assessment tool for teachers and educators. By identifying strengths and weaknesses it will allow	The consideration of this comment is located at different places in the text. These are highlighted in red in the lines: - 130 to 135 (page 7) - 550 to 563 (page 24)

	for designed and targeted interventions and programs. We hope that this will help to start a momentum in France like Australians tools. - Secondly, our study provides further perspectives with a focus on PL measurement coupled with a measure of the quantity and quality of physical activity over time.	
3 Is the assumption that all other PL assessments will not work with the emerging adult? Is there any data suggesting other PL assessments will not work with this age group? Passport for Life (PHE Canada) does have a PL assessment designed for the 16 to 18- year-old age group. It should be noted that the Canadian Assessment for Physical Literacy (CAPL) is considered to have statistically significant reliability and validity does include measures of values and attitudes, as does the Passport for Life PL assessment.	Thank you for the comment and giving us the change to clarify We do not claim that the CAPL or PFL tools would not work for this age group but there is not yet the evidence of validity and reliability for this age group. In addition, the structure and functioning of these tools can be debated regarding the philosophical underpinnings of the PL concept (e.g., additional arithmetic logic). We have attempted to move closer to this philosophical grounding but also to use the new debates and advances to go beyond the previous tools. This comment is also in line with comment 25 of reviewer#2	The consideration of this comment is located at different places in the text. These are highlighted in blue in the lines: - 14 to 15 (page 2) - 64 to 71 (page 4-5) - 400 to 409 (page 18) - 418 (page 18) - 426 to 427 (page 19) - 438 to 440 (page 19) - 444 to 447 (page 19-20) - 452 to 453 (page 20) - 546 to 549 (page 24) - 587 to 589 (page 25)
4 Line 51 through 53 – you may want to include that the Long Term Athlete Development (LTAD) model (Canada) has been in practice since the mid-2000s. As well, there is an Aboriginal LTAD for the Indigenous population. Further, the LTAD	Thank you for your proposition. LTAD could be a good illustration of a cooperation and a convergence of actors. Nevertheless, it is not a specific model of physical literacy. We think this idea should be developed in depth if we want to use it especially because many authors	-

model influences all government funded athletic association, NGOs, and education.	(like Richard Bailley) have been quite critical about it. However, this would complicate our (already long!) discussion and we think that this is not the priority here. But this idea is very interesting for our further work, thank you.	
5 Line 70 – Again, other than developing a tool for this age group and expanding the PL assessment to include broader measures, what is the purpose of this PL assessment? For example, CAPL provides an analysis of data that informs government and NGOs on the health and well-being of Canadian youth.	Thank you for this comment. We have responded to this request by adding details on the use of the test: either for an empirical study and regulate teaching practices to guide them towards an improvement of all the dimensions of PL. Also, ELIP makes it possible to overcome certain limits of pre-existing tests. We have argued in this sense in the manuscript. This comment is also in line with your comment 1.	The consideration of this comment is located at different places in the text. These are highlighted in red in the lines: - 130 to 135 (page 7) - 550 to 563 (page 24)
6 Line 92 – I agree, there is an unfounded assumption that superior PL will lead to a healthy active lifestyle, again, wondering who is going to use the data from your assessment and for what purpose? The school, the individual, government or is your PL assessment intended to further the research and understanding of PL – perhaps a clear statement is needed.	Thank you. Now, we have better articulated the aim of the ELIP as per previous comments.	The consideration of this comment is located at different places in the text. These are highlighted in red in the lines: - 130 to 135 (page 7) - 550 to 563 (page 24)

7 Line 103 – I find this to be one of the most compelling aspects of your research and this paragraph could be expanded.	Thank you very much for this positive comment. In the new version of the manuscript, we have developed this paragraph by detailing the different philosophies to be considered. NB: we took the opportunity to change the order of some paragraphs that have been modified following your comments. In the same way, some titles allowed to bring more clarity.	The consideration of this comment is located at different places in the text. These are highlighted in Green in the lines: - 87 to 88 (page 5) - 91 to 95 (page 5-6) - 98 (page 6) - 101 to 102 (page 6) - 110 to 135 (page 6-7) - 521 to 526 (page 23) - 535 to 537 (page 23) - 541 to 546 (page 23-24) -
8 Line 119 – One of the concerns Robinson & Randall raised was the ease of use of many of the PL assessments at the time. If this tool is intended to be used by trained wellfunded researchers only then this not be considered. If the ELIP is to have broader use and appeal then along with the reliability and validity of the tool, one must consider the ease of use, e.g., can an untrained individual run the ELIP, does it require significant time?	Thank you for this pertinent comment. The ease of use is part of the feasibility announced in the introduction. The ELIP will be a research tool but also a practical tool for teachers and educators. On this point, we have added details. Statistical method will be necessary to reduce the test and get closer to the expectations of a practical tool Nevertheless, the major purpose of this article was to provide the foundation for the test. Future studies will aim to measure its implementation in Physical Education and its relevance to the development of teaching practices.	The consideration of this comment is located at different places in the text. These are highlighted in green in the lines: - 593 to 595 (page 25)
9 Line 149 – you note that the ELIP was developed in part for this specific age group as it is the last	Thank you, this data would be very interesting to support the choice of our	

time in their lives they are required to engage in physical education, did your group of experts include any individuals that have administered PL assessments in a school setting? If they did, it may be worth noting.	experts. Unfortunately, this is not data that we have available with precision. Some researchers have already worked with PL tests but none on the targeted age group since the specific tools do not exist. Moreover, we focused our study more on the philosophical foundations than on its implementation. Therefore, we think it is better not to add this idea.	
10 Line 377 – consider removing the word some.	Thank you. We deleted this word	
11 Line 392 – may want to reference Characteristics and conceptual framework of the easy play model (Lu & Steele, 2014).	Very interesting! We didn't know about this model; we have now added this reference.	The consideration of this comment is located in discussion. These are highlighted in red in the lines: - 416 to 418 (page 18)
12 Line 404 – sentence ending on 404 > Excellent!	Thank you! We appreciate.	
13 Line 437 – I very much like this statement, could be expanded.	We have tried to briefly elaborate on our comments to make this perspective clear.	The consideration of this comment is located in discussion. These are written in green in the lines: - 530 to 534 (page 23)
14 Line 450 – Excellent - as positioned this sets this assessment to directly inform PL development at all age groups.	Thank you!	
15 Page 678 – Star Challenge. Like other PL assessments, this test does not require the participant to use the accepted FMS throwing technique. What specific FMS domains are being	We have added more details about the tests in the discussion and especially discussed the specificity of the "Star Challenge" The test targets competencies	The consideration of this comment is located in discussion. These are written in purple in the lines: - 462 to 476 (page 20-21)

evaluated in this assessment? Does the time factor influence the execution of technique during the assessment, does it produce anxiety that influences performance – what does the literature say. Very little discussion in the paper about the physical test.	beyond the FMS. A discussion has also been added to discuss time evaluation. Nevertheless, the Delphi method has put forward the tests, but the evaluation grids still need to be designed based on the pre-test results.	
16 It is not clear, why the bike challenge is included as a physical test within the PL assessment.	We believe that the resources required to be active for life encompasses bike and swimming. We have added paragraphs on the value and appropriateness of this experts' choices. Also, from the perspective of the Delphi method, the selection of experts is a valid argument. Nevertheless, this choice is also justified by the literature (e.g., Hulteen et al., 2018).	The consideration of this comment is located in discussion. These are highlighted in green in the lines: - 480 to 488 (page 21)
17 I highly commend that you have developed a tool that will assess the physical, cognitive, affective, and social domains of PL.	It is a pleasure to hear your positive feedback!	

Response to reviewer 2

Reviewer's comments	Response to reviewer	Effective changes
18 This research aims to contribute to the ongoing development of assessment for PL by developing an assessment tool for 'emerging adults. While I applaud the authors for identifying this gap in the literature and for their rigorous methodology approach a few things need addressing:	Thank you for your time and for this positive comment. We have responded to all your constructive comments, which helped to improve our manuscript.	The consideration of each comment is located at different places in the text with different colors. We have also specified the page and the line in this column. Regarding our methodology and in accordance with your comment, we have tried to highlight it by slightly modifying our title (line 2-3 page 1).

19 First, it is currently unclear as to which 'version' of PL the authors are adopting to inform the ELIP. The paper would benefit from a clearer positioning of PL – as it currently read there are many contradictions. For example, within the introduction, Whitehead (2001, 2007), Cairney et al. (2019), Keegan et al. (2019) and the IPLA (2017) are all cited/discussed despite them each offering different versions/definitions of PL. The authors appear to be conflating these different versions of PL into one. Transparency is needed to help support the validity and reliability of the ELIP.	Our introduction aimed to expose that the different definitions of PL highlight different dimensions but also have points of convergence, notably its holistic aspect and its focus on promoting PA for life. A Delphi method was used to design an evaluation tool based on different points of view. We tried to observe what elements would be the consensus within this expert diversity. Our experts were deliberately drawn from multiple different perspectives though, not all IPLA or not all LTAD for example – we think that is a strength of the study. Following your comment, - we clarified the use of the different definitions in the introduction. - we have clarified our initial objective, which was to build a tool based on a consensus of experts with different opinions of PL. That is why we do not anchor the ELIP in only one definition in the introduction– the tool is intended to cater for researchers from different perspectives. In view of the results obtained, your comment has made us aware that it would be interesting to place ELIP in the definition that is most in line with the consensus position used in this study. We	The consideration of this comment is located at different places. These are highlighted in yellow in the lines: - 33 to 36 (page 3) - 41 (page 3) - 46 to 47 (page 4) - 210 to 213 (page 10) - 511 to 519 (page 22)

	have discussed this and taken a position on it in the discussion.	
20 Secondly, the authors state that 'there is a discrepancy between the philosophical basis for the definition of PL' however it is unclear which philosophical basic and definition the authors are talking about? Whitehead? This needs to be more explicit, especially given that Whitehead has stated that it is inappropriate to assess PL, and rather proposes the 'charting' of an individual's PL journey. I suggest the authors engage with chapter six of Whitehead (2019) book. This is also important given that one of the stated goals of developing the ELIP is to 'adequately captured main philosophical underpinnings of PL' despite Whitehead/the IPLA already doing so. Further, if the ELIP is informed by phenomenology, existentialism, and monism then discussion and clarification is needed.	Thank you for this feedback and constructive comment which allowed us to clarify and improve our manuscript. Indeed, we wrote about the philosophical basis of Whitehead, but it was not clear enough. Some precisions were brought into introduction regarding the philosophical basis: including monism, existentialism, and phenomenology perspectives. Also, the question of the inappropriateness of the PL assessment was introduced and discussed. This "against" view was introduced and discussed in relation to the "agree" views. In our opinion, this question also goes beyond the construction of the tool since it is dependent on its future use. This major perspective was introduced in our manuscript and will certainly have to be developed in our further work. Finally, the discussion was also modified on this aspect. A discussion on the anchoring of ELIP in each of the philosophical perspectives was added. This comment is also in line with comment 7 of reviewer#1.	The consideration of this comment is located at different places in the text. These are highlighted in Green in the lines: - 87 to 88 (page 5) - 91 to 95 (page 5-6) - 98 (page 6) - 101 to 102 (page 6) - 110 to 135 (page 6-7) - 521 to 526 (page 23) - 535 to 537 (page 23) - 541 to 546 (page 23-24)

	NB: we took the opportunity to change the order of some paragraphs that have been modified following your comments. In the same way, some titles allowed to bring more clarity.	
21 Page 2, Line 45 states that 'PL experts agree on the holistic perspective' but this is not a case. In the line above Edwards et al. (2017) is cited – I believe the authors would benefit from a deeper analysis of this paper. Edwards et al highlight that not all PL assessments align with the philosophical underpinnings of Whitehead's version of PL. The separation of the physical, cognitive, social and affective domains in the ELIP also don't speak to holism.	This point is indeed ambiguous, so, based on your feedback and comment, we have modified our statement by adding "most of the PL experts". But we believe that the real experts agree on this aspect of a holistic PL. The European Erasmus+ PL project (PL4L) which put together researchers from both idealistic and pragmatic perspective emphasized the holistic perpective of PL. This is also the case for most associations dealing with physical literacy (in particular IPLA, Australian, Play for life, ...) Our comments are supported by Whitehead who stated in 2019: "*while modified definitions can cause confusion in a number of ways, particularly in relation to implications for practice and charting progress, a survey of most current definitions shows that, broadly, two essential aspects of the concept are retained. These are: a determination to promote commitment to PA for life; and the appreciation of the holistic nature of*	The consideration of this comment is located line 32 (page 3). The second comment has been considered in green with comment 7 and 20: - 87 to 88 (page 5) - 91 to 95 (page 5-6) - 98 (page 6) - 101 to 102 (page 6) - 110 to 135 (page 6-7) - 521 to 526 (page 23) - 535 to 537 (page 23) - 541 to 546 (page 23-24)

	human beings that must be recognized and addressed if this first commitment is to be realized". (Whitehead, 2019, p.10). Furthermore, the 2017 literature review (Edwards et al., 2017) pointed out that at least half of the definitions meet these holistic perspectives. Thus, we are convinced that a reductive vision can no longer be considered as representative of the PL concept. Your second comment is interesting, and offers the opportunity to provide some precisions about what we intend to do next. Experts all agreed on the existence of 4 dimension. To respect the holistic and monistic perspective, the data extracted should be treated according to complex system theory, that is, by considering dimensions and their respective interactions. In that we do not seek to adopt a reductionist perspective that "scores" physical literacy by adding the dimension in a linear fashion, but rather identify and characterize patterns of practitioners using nonlinear methods. This question deserves to be discussed further to invite future research on the issue. We have added elements in the introduction and discussion sections to	

	clearly inform the reader about this question (in line with comment 7 and 20).	
22 Figure 1 could do with more detail – is there a way the information from the appendices could be integrated into the figure for easier readability?	We change the Figure 1 according to your first comment. Nevertheless, we tried to integrate information from the other appendices as you suggested, but the result is less visual. We hope that our changes helped outline the details.	Please, see line 902
23 The ELIP would also benefit from further discussion regarding the chosen assessment items. For example, why was a bike and aquatics competence test chosen for the physical domain? Do a lot of emerging adults in France ride bikes and swim?	We have added some discussion, especially on these two tests and on the physical activities favored by emerged adults. Moreover, these ideas were strongly emphasized by our expert panels. From the perspective of the Delphi method, this is a valid argument.	The consideration of this comment is located in discussion. These are highlighted in green in the lines: - 480 to 488 (page 21)
24 No substitutions are offered for those who cannot swim and/or ride a bike.	This is indeed an issue that has given us a lot of thought. The ELIP provides an assessment of resources that have been deemed representative of the attributes of PL. The purpose of the ELIP is not to pose a certifying assessment, but to determine to what extent the participants were able to grasp most opportunities to be physically active. Walking is natural to human; this is not the case for biking and swimming which are culturally rooted. Being competent in those activities opens a range of perspective for PA (leisure, competition and even active transportation for biking) which were considered by our panel expert as pivotal to increase opportunities for PA.	The consideration of this comment is located at different places in the text. These are highlighted in red in the lines: - 130 to 135 (page 7) - 550 to 563 (page 24)

	Therefore, an individual who cannot swim or ride a bike should not be able to substitute for an assessment of these resources. ELIP would help him/her and teachers to be aware of vulnerability. This information is crucial for teachers/educators. We hope that by providing detailed information about the usefulness of the test, the justification for biking and swimming as key assets to grasp more opportunity for active lifestyle will become clearer. Thus, and agreeing with Reviewer#1's comment and your following comment, we have added this information about the usefulness of the test.	
25 With so many PL assessments on offer I believe a more detailed discussion is needed regarding 'what is it about the ELIP that differentiates it/makes it better from other PL assessments?'	We have added information about the philosophical underpinnings of the ELIP in response to your comment and this is also in line with comment 3. We believe that this provides additional insight into the differentiation of the ELIP from other tests. We also add more discussion on the difference with other tests and the specificity of ELIP. But rather than a real point-by-point comparison of the tools to show that the	The consideration of this comment is located at different places in the text. These are highlighted in blue in the lines: - 14 to 15 (page 2) - 64 to 71 (page 4-5) - 400 to 409 (page 18) - 418 (page 18) - 426 to 427 (page 19) - 438 to 440 (page 19) - 444 to 447 (page 19-20) - 452 to 453 (page 20) - 546 to 549 (page 24) - 587 to 589 (page 25)

	ELIP is “better” than the other tests, we think that this discussion should be carried out throughout the discussion, emphasizing the originality of the ELIP and the complementary view that it can bring to what already exists. NB: we took the opportunity of your comment to add a very recent measurement tool to our comments (the PPLAQ, Mota et al., 2021).	
26 You state towards the end of the discussion that the ELIP might help teachers’ guide and support students in their PL journey, but no guidance is offered to teachers regarding how to do so.	We agree with your remark and added more details on this points. We hope that it will help teachers to adjust their programs identifying the strong and weak points of their students. Nevertheless, studies are needed to verify that this tool has a pedagogical and not only a diagnostic purpose. We have added details on this question.	The consideration of this comment is located in discussion. These are highlighted in pink in the lines: - 556 to 557 (page 24)
27 Why was PubMed selected as the database from which to recruit PL ‘experts’?	PubMed is one of the major scientific database internationally recognized, which is why it was a natural choice to as. Also, the number of matches with the keyword “physical literacy”" have exponentially increased over the last two decades this database. This allowed us to obtain a base of experts who were solicited to suggest other experts (i.e. snowball sampling). We are aware of the limitations of this methodology; the one developed by Chen et al (2020) is more robust due to using “”CiteSpace”(v. 4.0.R5 SE,	-

	https://sourceforge.net/projects/citespace/) We almost definitely missed some experts – as would many similar studies. However, this allowed us to recruit a panel of 30 (<15) experts who met the criteria for a quality Delphi Method (COSMIN)	
28 Page 2 line 8 there appears to be an 'in' missing from this sentence: "designed to help reduce tensions IN existing approaches to PL assessment"	We have modified it.	**Line 8 to 9:** that has been designed to help reduce tensions existing in approaches to PL assessment resulting in a low uptake in applied settings
29 Page 1 Line 37 –Starting this paragraph off with the following quote from Bailey (2020) "PL is a promiscuous concept, applied in a wide range of settings, with different aims, approaches, and audiences" feels random and out of place. No additional text is given linking this sentence to the paragraph. Quotations should be used to reinforce your thoughts, ideas and claims.	We completely agree with your comment. We therefore deleted the quote.	

940