








## Article

# Validation of the 15-Item and 5-Item Versions of the Perceived Physical Literacy Instrument for Spanish Adolescents Aged 11–18: A Study Using the Original 18-Item Version

José Antonio Romero-Macarrilla <sup>1</sup>, Robert Bauer <sup>2,\*</sup> , Javier Fernández-Sánchez <sup>2,\*</sup> , Eva Fernández-Sánchez <sup>2</sup>, Iván González-Gutiérrez <sup>3</sup>, José Carmelo Adsuar <sup>4</sup> , Raquel Pastor-Cisneros <sup>5</sup> , María Mendoza-Muñoz <sup>6</sup> , Jorge Carlos-Vivas <sup>5</sup>  and Daniel Collado-Mateo <sup>2,\*</sup> 

<sup>1</sup> Faculty of Teaching Training, University of Extremadura, Avenida de la Universidad, S/N, 10071 Cáceres, Spain; jaromerom01@unex.es

<sup>2</sup> Research Centre in Sports Science (CIDE), Rey Juan Carlos University, 28942 Fuenlabrada, Spain; eva.fernandezsa@urjc.es

<sup>3</sup> Faculty of Social and Human Sciences, Universidad Europea del Atlántico, 39011 Santander, Spain; ivan.gutierrez@uneatlantico.es

<sup>4</sup> BioErgon Research Group, Faculty of Sport Sciences, University of Extremadura, 10003 Cáceres, Spain; jadssal@unex.es

<sup>5</sup> Physical Activity for Education, Performance and Health (PAEPH) Research Group, Faculty of Sport Sciences, University of Extremadura, 10003 Cáceres, Spain; raquelpc@unex.es (R.P.-C.); jorgecv@unex.es (J.C.-V.)

<sup>6</sup> Department of Communication and Education, Universidad Loyola Andalucía, 41704 Seville, Spain; mamendozam@unex.es

\* Correspondence: robert.bauer@urjc.es (R.B.); javier.fernandezsa@urjc.es (J.F.-S.); daniel.collado@urjc.es (D.C.-M.)

## Abstract

**Background:** Physical literacy is a multidimensional construct encompassing physical competence, confidence, motivation, knowledge, and lifelong engagement in physical activity. The Perceived Physical Literacy Instrument (PPLI) has been widely used internationally; however, previous adolescent validations have been based on a reduced 9-item version originally developed for teachers. This study aims to evaluate the validity and test–retest reliability of a Spanish adaptation of the original 18-item PPLI in Spanish adolescents aged 11–18 years. **Methods:** A multi-phase validation study was conducted with 869 Spanish adolescents (421 females). The procedure included: (1) translation and cultural adaptation, (2) Exploratory Factor Analysis (EFA;  $n = 290$ ), Confirmatory Factor Analysis (CFA;  $n = 579$ ) and invariance analyses, and (3) test–retest reliability assessment. **Results:** EFA supported a three-factor solution comprising 15 items. CFA showed standardized factor loadings ranging from 0.62 to 0.89, indicating that the latent constructs were adequately represented. Although the 15-item model showed acceptable fit, a 5-item unidimensional short form was developed due to limitations in the three-dimensional models. This short form demonstrated good model fit (scaled RMSEA = 0.073; scaled CFI = 0.992; SRMR = 0.026), adequate convergent validity (AVE = 0.558), high reliability ( $\omega = 0.821$ ), moderate test–retest stability (ICC = 0.69), and full configural, metric, and scalar longitudinal invariance. **Conclusions:** The 15-, 9-, and 5-item versions of the PPLI are valid and reliable options. The 15-item version allows comprehensive assessment and domain-level interpretation. The 9-item version facilitates comparability with previous international research. The 5-item version may be useful in contexts with time constraints but may not be the preferred choice for comprehensive assessment of physical literacy in clinical or detailed pedagogical diagnostic settings.



Academic Editor: Antonios E. Koutelidakis

Received: 3 March 2026

Revised: 30 March 2026

Accepted: 7 April 2026

Published: 9 April 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

**Keywords:** physical literacy; physical activity; physical education; public health; sedentarism; psychometrics

---

## 1. Introduction

Physical literacy (PL) has emerged in the last two decades as a core construct in physical education, sport, and public health. Based on Whitehead's philosophical framework, PL is commonly defined as the motivation, confidence, physical competence, knowledge, and understanding that enable individuals to value and take responsibility for lifelong engagement in physical activity [1,2]. PL has several domains, such as physical, cognitive, affective and social dimensions, and involves a holistic journey throughout the life course with special relevance in education and health [3]. In line with this conceptualization, national and international frameworks such as WHO Global Action Plan on Physical Activity 2018–2030, UNESCO's Guidelines on Quality PE, and the current Spanish education law (LOMLOE) have positioned PL as a central educational aim for promoting lifelong active and healthy lifestyles [4–6].

Adolescence is a critical developmental period for PL. More than 80% of adolescents aged 11–17 years do not meet physical activity recommendations. This phenomenon has been relatively stable in the last decade [7] and is especially worrying in the case of female adolescents, with more than 84% not meeting the recommendations [8]. Higher levels of PL have been associated with enhanced cardiorespiratory fitness and higher physical activity levels [9,10]. This suggests that improving PL may enhance physical activity and physical well-being in children and adolescents. In line with this hypothesis, a recent meta-synthesis study explored the effects of PL interventions on several crucial variables related to engagement in lifelong physical activity [11]. Specifically, the authors observed improvements in affective and psychological capabilities, including enhanced enjoyment, self-awareness, confidence, motivation, resilience, and self-worth; social capabilities, with evidence of increased engagement, collaboration, leadership, improved behavior, strengthened peer relationships, and greater social interaction; physical/motor capabilities, involving fundamental movement skills, coordination, object manipulation, and sport-related competence; and cognitive outcomes such as increased knowledge and awareness of physical activity, enhanced problem-solving, strategy and planning skills, improved focus and tactical reasoning, and greater body awareness.

PL is extremely important in childhood and adolescence because sedentary behaviors often increase during the transition to young adulthood [12], which may be a major public health challenge. PL might be a potential determinant of health across the life span, since it represents a multidimensional, reciprocal engagement cycle integrating motor competence, motivation, positive affect, social processes, and knowledge [13]. Thus, PL may promote sustained participation in physical activity, which in turn induces physiological, psychological and social adaptations that are associated with reduced risk of chronic disease and enhanced wellbeing. From this framework, higher PL in childhood and adolescence may contribute to more active habits and healthier profiles in adulthood.

As a multidimensional variable, PL may be divided into several domains or components. In the review and meta-analysis by Jiang et al. [10], four domains were identified based on the scientific articles included: (a) physical competence, (b) daily behavior, (c) knowledge and understanding, and (d) motivation and confidence [10]. However, other domains such as affective/psychological capabilities, social capabilities, physical/motor capabilities, and cognitive capabilities [11] have been suggested. This conceptual multidimensionality, together with the context-dependent nature of PL, makes valid assessment

methodologically challenging. A recent systematic review identified a range of PL assessment tools and concluded that most available instruments focus narrowly on either fundamental movement skills or physical fitness, while only a minority of tools assess affective, cognitive and physical components in an integrated way [14].

Among the developed tools to evaluate PL, the Perceived Physical Literacy Instrument (PPLI) is one of the most widely used self-report scales. Sum et al. [15] originally created an 18-item instrument for physical education teachers, derived from an extensive literature review, focus groups with experienced teachers, and expert panel review [15]. Exploratory and confirmatory factor analysis in a sample of Hong Kong physical education teachers yielded a three-factor structure that included (a) sense of self and self-confidence, (b) self-expression and communication with others, and (c) knowledge and understanding. The final version comprised 9 items with satisfactory reliability and model fit. Subsequently, the same 9-item tool was validated with a large sample of Hong Kong adolescents aged 11–19 years, showing good factorial validity, convergent and discriminant validity, and measurement invariance across gender [16]. In this study, the authors emphasized that the item wording was “generic” and not tied to a specific profession, which facilitated use of the teacher-based items with a sample comprising adolescents without further qualitative redevelopment.

The 9-item PPLI has been translated and validated in several countries and across different age groups. In Spain, Mendoza-Muñoz et al. [17] validated a Spanish version for adults derived from the original 18-item version of the PPLI. The authors conducted both an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA) and reported a valid tool version with good internal consistency, involving 9 items grouped into three dimensions: (1) physical competence, (2) motivation and confidence, and (3) knowledge and understanding. Similarly, López-Gil et al. [18], performed a cross-cultural adaptation and psychometric validation of the 9-item version (S-PPLI) in a sample of 360 Spanish adolescents aged 12–17 years. In this case, the authors used the reduced 9-item version validated in Hong Kong adolescents and conducted only a CFA with the Spanish sample. Their results supported the three-factor structure, with acceptable internal consistency, moderate–good test–retest reliability, and adequate convergent and discriminant validity. However, from both theoretical and measurement perspectives, this validated version has several relevant limitations. First, the item reduction from 18 to 9 items was conducted exclusively with a sample of physical education teachers, using their responses to decide which items to retain or discard [15]. It is therefore unclear whether the items that best discriminated among experienced adult professionals are also optimal for adolescents’ perceptions, given their different life contexts, responsibilities, and needs. Moreover, the 9 retained items represent only part of the original content domain. For instance, items concerning establishing friendships through sport or turning sport into an ongoing life habit were removed during the teacher-based psychometric trimming. Thus, based on the perspective of Whitehead’s framework [2,19] and recent pedagogical models of PL in physical education, these omitted elements may be highly relevant to assess a holistic construct such as PL. In addition, reviews of PL assessment have warned that many instruments, including frequently used questionnaires, risk oversimplifying the construct, over-emphasizing selected components (often physical competence), and under-representing its dispositional and lifelong character [14].

Given these considerations, the present study aimed to evaluate the validity and test–retest reliability of a Spanish adaptation of the original 18-item PPLI in adolescents aged 11–18 years. By retaining all original items, we aimed to (1) consider the full holistic scope of the PL construct as originally conceptualized through extensive literature review, focus groups, and expert panel review [15]; (2) analyze which items best capture perceived

PL in the Spanish adolescent context, potentially different from the teacher-based item selection; (3) generate new Spanish items and psychometric data that reflect adolescent perspectives and realities; and (4) create a more comprehensive measurement tool, as well as to explore the possibility of developing a short form for some context where the number of items and the response time are decisive.

## 2. Materials and Methods

### 2.1. Design

A multi-phase validation study employing exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and test–retest reliability assessment was conducted. The research was reviewed and approved by the Bioethics and Biosafety Committee of the University of Extremadura and is in line with Spanish research ethics guidelines and the updated Declaration of Helsinki [20]. All participants provided written informed consent, and adolescents provided written assent. All procedures were conducted with the consent of the parents or legal guardians of the adolescents.

### 2.2. Participants

The sample consisted of 869 adolescents (421 females) aged 11–18 who completed the questionnaire once. A subsample involving 106 adolescents completed the questionnaire again two months later to assess test–retest reliability. Participants were recruited from 16 secondary schools located in different regions of Spain (north, south, west, and east). Of these, 13 were public schools, two were semi-private (government-funded) schools, and one was a private institution. Regarding the geographical context, two schools were located in urban areas with more than 100,000 inhabitants, five in urban areas with populations between 50,000 and 100,000, four in intermediate areas (10,000–50,000 inhabitants), and five in rural areas with fewer than 10,000 inhabitants.

Participants were eligible if they met the following criteria: (a) were adolescents aged 11–18, (b) were enrolled in a secondary education institution at the time of data collection, (c) were able to read and write in Spanish to understand and complete the questionnaires autonomously, (d) provided informed consent, and (e) had parental or legal guardian authorization.

### 2.3. Instrument

The Perceived Physical Literacy Instrument (PPLI) is a self-report questionnaire based on the holistic Whitehead's conceptualization of PL involving confidence, competence, knowledge, understanding, and motivation for lifelong engagement in physical activity. The original PPLI was developed for physical education teachers in Hong Kong through literature review, focus groups, and expert consultation, and subsequently refined to a 9-item version with a three-factor structure supported by exploratory and confirmatory factor analyses [15]. The same 9-item version was later validated in adolescents from Hong Kong, confirming adequate factorial validity and measurement invariance across sex [16]. Items from PPLI are rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Items are equally distributed across three dimensions: (1) sense of self and self-confidence, (2) self-expression and communication with others, and (3) knowledge and understanding. Subscale scores are calculated by summing the three items within each factor, and a total score is obtained by summing all items, with higher scores indicating greater perceived physical literacy. The Spanish adolescent version (S-PPLI) was validated through CFA from the refined 9-item version validated for teachers in Hong Kong. It demonstrated good model fit in confirmatory factor analysis (CFI = 0.976; RMSEA = 0.057; SRMR = 0.031), satisfactory internal consistency (Cronbach's  $\alpha \approx 0.87$ ; Mc-

Donald's  $\omega \approx 0.87$ ), and moderate-to-good test–retest reliability (ICC = 0.62–0.79). Despite its validation, the 9-item version of the Spanish version presents important theoretical and measurement limitations detailed in the Section 1. The main limitation is related to the reduction from 18 to 9 items based on responses from physical education teachers in Hong Kong, raising concerns about whether the retained items are appropriate for Spanish adolescents.

#### 2.4. Procedure

To develop and assess the validity and reliability of the PPLI in Spanish adolescents, a cross-cultural adaptation process was conducted in accordance with well-known guidelines for the adaptation of self-report measures [21].

The procedure included three phases: (1) translation and cultural adaptation of the original 18-item PPLI [15], (2) EFA, CFA and invariance analyses for Spanish adolescents, and (3) test–retest reliability of the PPLI.

##### 2.4.1. Phase 1: Translation and Cultural Adaptation to Develop the Spanish Version of PPLI for Adolescents

This phase involved five stages beginning with translation of the original 18-item version developed by Sum et al. [15]. This was conducted by two independent bilingual translators who are fluent in Spanish and English and with previous experience in scientific translation [22]. The two translators separately translated the original English version into Spanish.

In the second stage, the two preliminary translations were compared and combined into a single Spanish version. Discrepancies were discussed until agreement was reached. This consensus version was subsequently evaluated by two members of the research team with academic expertise in the field of PL to ensure semantic accuracy, conceptual equivalence, and clarity of wording, while preserving the intent of the original instrument [23,24].

The third stage was the back translation. Here, the consolidated Spanish version was back-translated into English by an independent bilingual translator who was blinded to the original questionnaire. This step was conducted to verify conceptual fidelity and to detect potential inconsistencies or deviations in meaning [25].

After the back translation, in the fourth stage, an expert committee composed of three specialists with scientific expertise and professional experience working with adolescents reviewed and provided feedback and recommendations based on all generated materials, including the original instrument, forward translations, consensus version, and back-translation. In this stage, a pre-final version was obtained, ready to be tested with the target sample.

In the fifth and final stage, the version developed in the previous stage was tested with a sample of 30 adolescent students to evaluate clarity, comprehensibility, and cultural appropriateness. Based on the feedback provided by these 30 participants, no further modifications were needed. The resulting version was therefore considered suitable for subsequent psychometric evaluation in the target population.

##### 2.4.2. Phase 2: Exploratory (EFA) and Confirmatory (CFA) Factorial Analyses

The final Spanish version of the PPLI developed in the Phase 1 was administered using QR codes or direct links to a Google Forms survey in high schools located in different regions of Spain. Before data collection, parents or legal guardians were informed about the study through detailed information sheets disseminated via the schools' digital learning platforms, where an opt-out form was available. Student involvement was entirely voluntary, and all participants provided informed consent after being clearly told that choosing not to participate would have no impact on their academic standing or evaluation.

The original sample ( $N = 869$ ) was randomly divided into two subsamples in a 2:1 ratio. Exploratory factor analyses (EFAs) were performed with the first subsample ( $n = 290$ ), and confirmatory factor analyses (CFAs) with the second subsample ( $n = 579$ ), following best-practice recommendations for scale development and cross-validation [26–28]. The characteristics of the sample are shown in Table 1.

**Table 1.** Sociodemographic Data.

Sociodemographic Variable	n	%
Gender		
Female	421	48.45%
Male	448	51.55%
Age		
11	4	0.46%
12	133	15.30%
13	124	14.27%
14	128	14.73%
15	197	22.67%
16	169	19.45%
17	102	11.74%
18	12	1.38%

$n$  = sample size; % = percentage.

#### 2.4.3. Phase 3: Test–Retest Reliability

From the original sample ( $N = 869$ ), a subsample involving 106 adolescents participated in this phase. They answered the form a second time two months after the first data collection. Again, parents or legal guardians were informed through the schools' digital learning platforms, and an opt-out form was available. Similarly, student involvement was entirely voluntary, and all participants provided informed consent and were told that there would be no negative consequence if they decided not to participate. The mean age of this subsample was 13.45 and the SD was 1.35. The gender distribution was 1:1, involving 53 male adolescents and 53 female adolescents.

#### 2.5. Statistical Analysis

Statistical analyses were conducted in R (version 4.4.2; [29]) using RStudio (version 2025.05.0+496; [30]). Exploratory and confirmatory factor analyses were performed using the lavaan (version 0.6-19) [31,32] and semTools (version 0.5-7) [33] packages.

The full sample was randomly split into two subsamples. Exploratory factor analyses (EFAs) were conducted on the first subsample to identify the underlying factor structure. Confirmatory factor analyses (CFAs) were subsequently conducted on the second subsample to evaluate the fit of the EFA-derived models and to compare them with theoretically derived alternative models.

The chi-square test evaluates exact model fit and is therefore highly sensitive to sample size. As sample size increases, even minor deviations between the observed and model-implied covariance matrices are likely to yield statistically significant chi-square values [23,34]. Model evaluation relied primarily on approximate fit indices, including the root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), Tucker–Lewis index (TLI), and standardized root mean square residual (SRMR). These indices provide a more informative assessment of practical model adequacy in large samples [28,35,36].

### 2.5.1. Exploratory Factor Analyses

An EFA was conducted on the 18 items using data from 290 participants. Prior to extraction, sampling adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s test of sphericity. The number of factors was determined using parallel analysis. Factors were extracted using the robust weighted least squares estimation (WLSMV) with oblique (Promax) rotation, allowing for correlated factors [37–39].

Items were considered for removal if they showed one or more of the following problems: (a) no salient loading (all loadings < 0.40), (b) strong cross-loadings ( $\geq 0.40$  on two or more factors or small differences in loadings), (c) low communality ( $R^2 < 0.30$ ), or (d) conceptual misfit with the intended construct. After item refinement, factor structure and model adequacy were re-evaluated using commonly recommended criteria, including scaled RMSEA, scaled CFI, scaled TLI, and SRMR [35,40]. As the model was estimated as an exploratory factor model within the SEM framework using lavaan [31,32], model fit indices correspond to SEM-based fit statistics.

### 2.5.2. Confirmatory Factor Analyses

CFAs were conducted using data from the second subsample ( $n = 579$ ). Given the ordinal nature of the Likert-type scale items, the models were estimated using the robust weighted least squares estimator (WLSMV). Robust (scaled) model fit indices were assessed, including robust RMSEA, robust CFI, robust TLI, and SRMR [35,36,40,41]. Convergent validity was evaluated using the Average Variance Extracted (AVE) values and the magnitude of standardized factor loadings [42]. Internal consistency reliability was assessed using McDonald’s  $\omega$ , which provides a model-based estimate of composite reliability suitable for latent variable models [43,44]. In addition to EFA-derived models, previously published model structures [15,16] were tested for comparison.

### 2.5.3. Test–Retest Reliability

Test–retest reliability was examined using intraclass correlation coefficients (ICCs) based on latent factor scores. CFAs were estimated separately for Time 1 and Time 2, and factor scores were extracted using empirical Bayes estimation. ICCs were calculated using a two-way random-effects model assessing consistency for single measurements [ICC(C,1)], treating the two measurement occasions as “raters.” This approach evaluates the temporal stability of individual differences while allowing for potential mean-level changes over time [45,46].

ICCs were computed for the unidimensional model structures. No ICCs were reported for the second-order factor models, since ICCs assume that the same latent variable is measured on an equivalent scale across measurement occasions, such that individual differences can be meaningfully compared over time [45]. In second-order CFA models, however, higher-order factors do not have unique indicators and are instead inferred indirectly through first-order latent variables. Factor scores for second-order constructs are highly dependent on model scaling and estimation at each time point and are not expected to show stable rank ordering when models are estimated separately [46,47]. Accordingly, ICCs are not an appropriate reliability index for second-order latent constructs. For this reason, test–retest reliability was evaluated only for unidimensional models, for which latent factor scores are well defined and directly comparable across time.

### 2.5.4. Longitudinal Measurement Invariance

Longitudinal measurement invariance was evaluated for models demonstrating adequate test–retest reliability. Invariance testing followed a hierarchical sequence of increasingly restrictive models, including configural invariance (same factor structure across time),

metric invariance (equal factor loadings), and scalar invariance (equal factor loadings and intercepts). Model comparisons were conducted using scaled chi-square difference tests.

Because sparse response categories in some ordinal items prevented stable estimation under WLSMV, invariance testing was conducted using a continuous approximation with robust maximum likelihood estimation (MLR), a commonly applied pragmatic solution when strict categorical invariance testing is infeasible [28,36,48]. Evidence for invariance was inferred from non-significant chi-square difference tests between nested models. Furthermore, measurement invariance testing was not pursued for models exhibiting insufficient test–retest reliability or estimation problems, as such models do not permit meaningful longitudinal comparisons.

Due to the potential differences between younger and older adolescents, we also examined measurement invariance of the three models to test and compare whether the scale was equally appropriate for age groups. Multi-group confirmatory factor analyses were used. Age groups were defined as early adolescence (11–14 years) and middle-to-late adolescence (15–18 years), following previous developmental research [49]. A progressive approach was applied, in which increasingly restrictive models were tested sequentially, including configural, metric, and scalar invariance. Configural invariance assessed whether the same factorial structure was present across groups. Metric invariance tested the equality of factor loadings across groups, indicating whether items contributed similarly to the latent constructs. Scalar invariance tested the equality of item thresholds, allowing for meaningful comparisons of latent means. Model comparisons were evaluated using changes in fit indices, with  $\Delta\text{CFI} \leq 0.01$  and  $\Delta\text{RMSEA} \leq 0.015$  indicating invariance. When full scalar invariance was not supported, partial scalar invariance was examined by freeing a minimal number of item thresholds based on modification indices.

### 3. Results

#### 3.1. Results of Exploratory Factor Analysis of the First Subset

As an initial assessment of sampling adequacy, the Kaiser–Meyer–Olkin (KMO) criterion indicated excellent suitability for factor analysis (KMO = 0.927). Bartlett’s test of sphericity was then conducted to evaluate the appropriateness of the data for factor analysis. The test was statistically significant,  $\chi^2(153) = 2545.98$ ,  $p < 0.001$ , indicating that the inter-item correlations were sufficiently strong to justify the application of exploratory factor analysis.

The optimal number of factors was determined primarily by Parallel Analysis, supported by the latent root criterion (eigenvalues  $> 1.0$ ), which clearly indicated a three-factor model. Nevertheless, this solution was compared with a two- and four-factor model, which showed poorer fit.

Of the original 18 items, three (Items 4, 5, and 17) showed no salient factor loadings ( $< 0.40$ ) and were removed. Inspection indicated that these items weakened factorial purity without adding unique variance. Their exclusion resulted in an improved model fit and a clearer simple structure, without compromising factor score adequacy. This solution was therefore retained for subsequent analyses.

After removing the problematic items, the three-factor structure was reconfirmed via a subsequent Parallel Analysis and eigenvalues. All retained items loaded  $\geq 0.40$  on a single factor, with minimal cross-loadings (see Table 2). Model fit was acceptable (scaled RMSEA = 0.067, 90% CI [0.053, 0.081]; scaled CFI = 0.981; scaled TLI = 0.969; SRMR = 0.035), based on commonly recommended criteria [35,40]. The three factors were moderately correlated ( $r = 0.65$ – $0.69$ ), supporting the use of an oblique rotation [37–39]. This model was then used in the CFA.

**Table 2.** Factor Structures of the Exploratory Factor Analysis (N = 290).

ID	Items (Analyzed Spanish Version)	Items (Original English Version)	F1	F2	F3	R <sup>2</sup>
Item 1	Se me dan bien las habilidades como correr, pillar, saltar o trepar.	I possess adequate fundamental movement skills.	<b>0.61</b>	0.10	0.09	0.57
Item 2	Para mi edad tengo buena forma física.	I am physically fit, in accordance with my age.	<b>0.99</b>	−0.06	−0.04	0.88
Item 3	Puedo aplicar las habilidades motrices que he aprendido a otras actividades físicas.	I am able to apply learnt motor skills to other physical activities.	<b>0.54</b>	0.11	0.22	0.62
Item 4 *	Tengo una actitud positiva e interés en los deportes.	I have a positive attitude and interest in sports.	0.29	0.19	0.39	0.62
Item 5 *	Me valoro o valoro a los/las demás cuando hacemos deporte.	I appreciate myself or others doing sports.	0.23	0.29	0.19	0.41
Item 6	Puedo aplicar los conocimientos aprendidos en educación física a largo plazo.	I am able to apply PE knowledge in the long run.	−0.14	<b>0.64</b>	0.16	0.44
Item 7	Soy capaz de gestionarme para mantenerme en buena forma física.	I possess self-management skills for fitness.	<b>0.63</b>	0.19	0.13	0.76
Item 8	Tengo las habilidades necesarias para evaluar mi estado de salud.	I possess self-evaluation skills for health.	<b>0.46</b>	0.35	0.10	0.67
Item 9	Hago deporte para mejorar mi salud.	I am willing to do sports for better health.	0.07	−0.04	<b>0.73</b>	0.57
Item 10	Puedo comunicarme bien a través de mi cuerpo.	I have strong communication skills.	0.00	<b>0.56</b>	0.29	0.63
Item 11	Tengo muy buenas habilidades sociales.	I have strong social skills.	−0.10	<b>0.66</b>	0.09	0.44
Item 12	Soy capaz de valerme por mí mismo/a en el entorno natural.	I am confident in wild/natural survival.	0.09	<b>0.86</b>	−0.22	0.60
Item 13	Puedo superar problemas y dificultades.	I am capable in handling problems and difficulties.	0.04	<b>0.86</b>	−0.29	0.50
Item 14	Tengo una mentalidad adecuada para realizar deporte a lo largo de toda mi vida.	I have a mindset for lifelong sports.	0.27	−0.02	<b>0.62</b>	0.66
Item 15	Considero que el deporte será un hábito para toda mi vida.	I can turn doing sports into an on-going habit of life.	0.11	−0.26	<b>0.98</b>	0.80
Item 16	Hago amistades gracias al deporte.	I establish friendship through sports.	−0.12	0.02	<b>0.83</b>	0.59
Item 17 *	Conozco los beneficios del deporte para la salud.	I am aware of the benefits of sports related to health.	0.00	0.28	0.35	0.35
Item 18	Me gusta estar al día de las nuevas tendencias deportivas.	I aspire to know the current sports trend.	−0.19	0.23	<b>0.64</b>	0.47

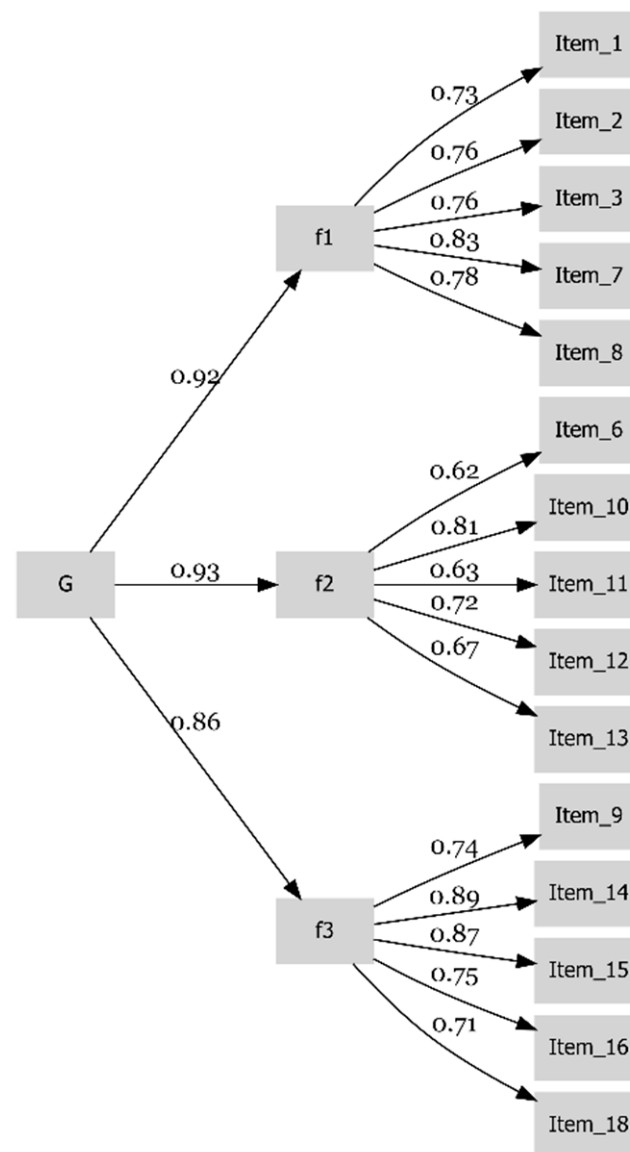
The presented values reflect the final model, except for removed items, whose values reflect the initial model (marked with \*); Factor loadings highlighted in bold indicate the respective main factors they load on; R<sup>2</sup> = Item Communality.

### 3.2. Results of the Confirmatory Factor Analyses of the Second Subset

A CFA was conducted to validate and confirm the three-factor structure of the PPLI scale found in the EFA. Figure 1 shows the standardized factor loadings of all items. These values ranged from 0.62 to 0.89, all exceeding the minimum threshold of 0.40 [37]. This indicates that the observed variables adequately represented the latent constructs.

Internal consistency was assessed using composite reliability, reported as McDonald's  $\omega$ , a model-based reliability coefficient estimated from the CFA model with ordinal indicators. Table 3 summarizes the convergent validity analyses. Factor 1 showed a McDonald's  $\omega$  of 0.848 and an AVE of 0.596, indicating adequate convergent validity for

the construct [39]. Factor 3 likewise demonstrated adequate convergent validity, with a McDonald's  $\omega$  of 0.851 and an AVE of 0.632. Although the AVE for Factor 2 was slightly below the recommended 0.50 threshold, composite reliability exceeded 0.70, indicating acceptable convergent validity [42].



**Figure 1.** Factor Loadings of the Three-Factor Model Structure with 15 Items for the Spanish Physical Literacy Instrument for Adolescents.

However, high covariance values between the three factors (0.86, 0.79, and 0.80) pointed toward a potential underlying G-factor that explained a large part of the scale's variance. This was confirmed by testing a bifactor model and a second-order model. The former failed to converge, while the latter showed high loadings of first-order factors (that is, 0.92, 0.93, and 0.86) on the second-order G-factor (see Figure 1). Model fit indices revealed a moderate-to-good fit between the data and the model (scaled RMSEA = 0.080, 90% CI [0.072, 0.087]; robust RMSEA = 0.099, 90% CI [0.090, 0.109]; scaled CFI = 0.967; robust CFI = 0.906; scaled TLI = 0.961; robust TLI = 0.887; SRMR = 0.048). The 15-item model with three first-order factors and a second-order G-factor described the data well.

We compared this model with the one found by Sum et al. [15,16]. Accordingly, we analyzed their reported 9-item model structure with our second dataset of 579 participants.

Similar to our model, we found a strong underlying second-order G-factor (with first-order factor loadings of 0.94, 0.85, and 0.88). Notably, model fit indices revealed a slightly better fit between the data and the model (scaled RMSEA = 0.075, 90% CI [0.060, 0.090]; robust RMSEA = 0.090, 90% CI [0.072, 0.109]; scaled CFI = 0.980; robust CFI = 0.950; scaled TLI = 0.970; robust TLI = 0.925; SRMR = 0.043) than the model found in our EFA.

With the different item sets showing a similar G-factor, together with the subscales explaining comparatively little unique variance in both models, there was strong evidence pointing toward an essentially unidimensional measure. In addition, as mentioned above, attempts to estimate a bifactor model further suggested the dominance of the general factor. Although convergence could be achieved under constrained specifications, the presence of negative residual variances (Heywood cases) indicated that the general factor absorbed nearly all item variance, leaving insufficient residual variance for specific factors. Such findings have been described as indicative of essential unidimensionality rather than evidence for substantively meaningful item-level multidimensionality [47,50]. Overall, these results suggested that, although multidimensional representations were plausible, the scale was primarily driven by a strong general factor.

**Table 3.** Results of the Convergent Validity Analyses for the 15-item version of the instrument.

Factor	Item	Std. Loading	$\omega$	AVE
Factor 1	Item 1	0.73	0.848	0.596
	Item 2	0.76		
	Item 3	0.76		
	Item 7	0.83		
	Item 8	0.78		
Factor 2	Item 6	0.62	0.780	0.481
	Item 10	0.81		
	Item 11	0.63		
	Item 12	0.72		
	Item 13	0.67		
Factor 3	Item 9	0.74	0.851	0.632
	Item 14	0.89		
	Item 15	0.87		
	Item 16	0.76		
	Item 18	0.71		
G-Factor	Factor 1	0.92		
	Factor 2	0.93		
	Factor 3	0.86		

Standardized factor loadings (Std.) are reported. Composite reliability is reported as McDonald’s  $\omega$ , estimated from the CFA model. McDonald’s  $\omega$  and Average Variance Extracted (AVE) are shown at the first-order factor level. The second-order G-factor is evaluated based on standardized loadings and overall model fit; McDonald’s  $\omega$  and AVE are not reported for higher-order factors.

### 3.3. Development of a Unidimensional Short Form

Given the dominance of the general factor and the goal of providing a parsimonious instrument for assessing the overall construct, a unidimensional short form was developed. Using the EFA subsample, a one-factor exploratory solution was estimated across all 18 original items. Items were evaluated based on their standardized factor loadings, communalities, item correlations, and conceptual relevance to the underlying construct.

Five items demonstrating the highest loadings and communalities, while maintaining adequate content coverage, were retained for the short form. This approach aligns with

established recommendations for theory-guided item reduction aimed at constructing reliable unidimensional scales [26,51]. The EFA showed good fit for this reduced set (scaled RMSEA = 0.057, 90% CI [0.000, 0.110]; scaled CFI = 0.996; scaled TLI = 0.992; SRMR = 0.025). Table 4 shows the model fit of the 5-, 9-, and 15-item versions of the instrument. Furthermore, correlations and differences between the models were explored, showing significant correlation coefficients higher than 0.85, which can be interpreted as very high. No significant differences were found among the three versions, and effect sizes were negligible (Cohen's  $d < 0.01$ ).

**Table 4.** Comparison of model fit indices across versions.

Model	Reduced 5-Item Version	9-Item Version by Sum et al. [15]	15-Item Version
Items (n)	5	9	15
RMSEA	0.057	0.075	0.080
CFI	0.996	0.980	0.967
TLI	0.992	0.970	0.961
SRMR	0.025	0.043	0.048
Standardized Mean (SD)	−0.02 (0.61)	−0.02 (0.64)	−0.01 (0.61)
Correlation With 15-Item Version	0.92	0.93	1.00
Correlation With 5-Item Version	1.00	0.87	0.92

RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; SRMR = Standardized Root Mean Square Residual. Scaled values are reported for RMSEA, CFI, and TLI. All correlations between the three model versions of the scale were significant ( $p < 0.001$ ). No significant differences (via pairwise  $t$ -tests;  $p > 0.05$ ) were found between the three versions of the scale. Cohen's  $d$  effect sizes were always lower than 0.01, which can be considered negligible.

The unidimensional five-item model was then tested using CFA in the independent validation subsample ( $n = 579$ ). The CFA demonstrated good overall fit to the data (scaled RMSEA = 0.073, 90% CI [0.042, 0.108]; robust RMSEA = 0.089, 90% CI [0.049, 0.132]; scaled CFI = 0.992; robust CFI = 0.981; scaled TLI = 0.984; robust TLI = 0.963; SRMR = 0.026). The robust RMSEA was elevated; however, RMSEA is known to be biased upward in models with small degrees of freedom and strong factor loadings, particularly in models with ordinal indicators [41,52]. Therefore, the pattern of fit indices supported the adequacy of the unidimensional model.

All standardized factor loadings were high ( $\lambda = 0.70$ – $0.79$ ), indicating strong associations between the latent factor and its indicators (see Table 5). Item  $R^2$  values ranged from 0.49 to 0.62, suggesting that a substantial proportion of variance in each item was explained by the latent construct.

**Table 5.** Results of the Convergent Validity Analysis of the 5-item version of the instrument.

ID	Items (Analyzed Spanish Version)	Items (Original English Version)	Std. Loading	$R^2$
Item 1	Se me dan bien las habilidades como correr, pillar, saltar o trepar.	I possess adequate fundamental movement skills.	0.70	0.49
Item 3	Puedo aplicar las habilidades motrices que he aprendido a otras actividades físicas.	I am able to apply learnt motor skills to other physical activities.	0.77	0.60
Item 4	Tengo una actitud positiva e interés en los deportes.	I have a positive attitude and interest in sports.	0.79	0.62
Item 8	Tengo las habilidades necesarias para evaluar mi estado de salud.	I possess self-evaluation skills for health.	0.70	0.49
Item 14	Tengo una mentalidad adecuada para realizar deporte a lo largo de toda mi vida.	I have a mindset for lifelong sports.	0.76	0.58

Standardized (Std.) factor loadings and  $R^2$  (Coefficient of Determination) estimates are reported. Composite reliability, reported as McDonald's  $\omega = 0.821$ , estimated from the CFA model. Average Variance Extracted (AVE) = 0.558.

The short form demonstrated high reliability, with a McDonald’s  $\omega$  of 0.821. The AVE was 0.558, exceeding the recommended threshold of 0.50 and indicating adequate convergent validity [42].

### 3.4. Model Comparisons

#### 3.4.1. Differences and Correlations Among Models

Table 6 shows the items included and the distribution in factors of the 5-, 9-, and 15-item versions of the instrument.

**Table 6.** Items Included in Each Version, and Differences and Correlations Between the Versions.

	Original English Item	15-Item	9-Item	5-Item
1	I possess adequate fundamental movement skills.	F1✓	✗	✓
2	I am physically fit, in accordance with my age.	F1✓	F2✓	✗
3	I am able to apply learnt motor skills to other physical activities.	F1✓	✗	✓
4	I have a positive attitude and interest in sports.	✗	F1✓	✓
5	I appreciate myself or others doing sports.	✗	F1✓	✗
6	I am able to apply PE knowledge in the long run.	F2✓	✗	✗
7	I possess self-management skills for fitness.	F1✓	F2✓	✗
8	I possess self-evaluation skills for health.	F1✓	F2✓	✓
9	I am willing to do sports for better health.	F3✓	✗	✗
10	I have strong communication skills.	F2✓	✗	✗
11	I have strong social skills.	F2✓	F3✓	✗
12	I am confident in wild/natural survival.	F2✓	F3✓	✗
13	I am capable in handling problems and difficulties.	F2✓	F3✓	✗
14	I have a mindset for lifelong sports.	F3✓	✗	✓
15	I can turn doing sports into an ongoing habit of life.	F3✓	✗	✗
16	I establish friendship through sports.	F3✓	✗	✗
17	I am aware of the benefits of sports related to health.	✗	F1✓	✗
18	I aspire to know the current sports trend.	F3✓	✗	✗

#### 3.4.2. Test–Retest Reliability (ICC)

Regarding the test–retest reliability, we compared three measurement models: (1) the short unidimensional model with five items, (2) the unidimensional model structure based on the original version by Sum et al. [15] with nine items, and (3) the unidimensional model structure derived from the present EFA with 15 items. For all models, CFAs were estimated separately at Time 1 and Time 2 using the WLSMV estimator. Test–retest reliability was assessed by extracting latent factor scores from each time point and computing ICCs (with two-way random effects, assessing consistency, single-measure), where the two measurement occasions were treated as the “raters.”

Test–retest reliability of the three models was very similar. All ICCs ranged from 0.689 to 0.707, SEM was between 0.35 and 0.37, and SRD was between 0.97 and 1.02 (see Table 7). Regarding the McDonald’s  $\omega$ , the best value was observed for the 15-item version (0.919), which was slightly higher than the one observed for the nine-item version developed by Sum et al. [15] and higher than the unidimensional model (0.821). Thus, all models achieved good temporal stability, suggesting that individual differences are preserved across measurement occasions to a meaningful extent. Both longer unidimensional models likewise demonstrated moderate to good test–retest reliability. However, the reduced five-item unidimensional model demonstrated the most favorable balance of parsimony, test–retest reliability, and longitudinal measurement invariance.

**Table 7.** Test–Retest Reliability and Longitudinal Measurement Invariance Across CFA Models.

Model	Items (n)	Scale Factor	ICC	95% CI	Longitudinal Invariance	SEM	SEM%	SRD	SRD%	$\omega$	AVE
Reduced 5-Item Version	5	Total score	0.690	[0.575, 0.778]	Configural, metric, scalar supported	0.35	8.78	0.97	24.34	0.821	0.558
9-Item Version by Sum et al. [15]	9	Total score	0.689	[0.575, 0.777]	Configural, metric, scalar supported	0.37	9.22	1.02	25.54	0.857	0.465
15-item version	15	Total score	0.707	[0.598, 0.791]	Configural, metric, scalar supported	0.35	8.80	0.98	24.40	0.919	0.503

ICC = intraclass correlation coefficient based on a two-way random-effects model assessing consistency [ICC(C,1)]; SEM = standard error of measurement; SRD = smallest real difference;  $\omega$  = composite reliability, reported as McDonald’s  $\omega$ , estimated from the CFA model; AVE = Average Variance Extracted. ICCs computed using latent factor scores derived from CFA models estimated separately at Time 1 and Time 2. Measurement invariance was tested longitudinally using multi-group CFA. Due to sparse response categories, invariance testing for ordinal indicators was conducted using a continuous approximation.

### 3.4.3. Measurement Invariance Across Time

Longitudinal measurement invariance was examined for the unidimensional models to determine whether the constructs were measured equivalently across Time 1 and Time 2 (see Table 7). Because some ordinal response categories were sparsely populated at one time point, invariance testing was conducted using a continuous approximation. Configural, metric, and scalar invariance models were estimated and compared using scaled chi-square difference tests.

The comparison between the configural and metric models was non-significant, indicating that factor loadings could be constrained to equality across time without a deterioration in model fit. Similarly, the comparison between the metric and scalar models was also non-significant, supporting the equality of intercepts across Time 1 and Time 2. These results provide evidence for at least scalar invariance of the unidimensional models over time, implying that observed changes (or stability) in factor scores can be interpreted as reflecting true change rather than measurement artifacts.

To facilitate interpretation of individual change scores, the standard error of measurement (SEM) and the smallest real difference (SRD) were calculated for the unidimensional models using the ICC-based reliability estimates. For the five-item model, the SEM was 0.35, corresponding to 8.8% of the total scale range. The SRD was 0.97, equivalent to 24.34% of the scale range, indicating the minimum change required to be confident that an observed difference exceeds measurement error at the individual level.

### 3.4.4. Measurement Invariance Across Age Groups

The configural model of the single-factor structure for the 5-item model demonstrated good fit (robust CFI = 0.972, robust RMSEA = 0.091, SRMR = 0.028), indicating a similar factorial structure across age groups. Metric invariance was supported, as constraining factor loadings resulted in negligible changes in model fit ( $\Delta$ CFI = 0.003;  $\Delta$ RMSEA =  $-0.018$ ). Scalar invariance was also supported across age groups. Constraining both factor loadings and item thresholds did not significantly worsen model fit ( $\Delta\chi^2(4) = 3.685, p = 0.450$ ), and changes in fit indices remained well below recommended cutoffs ( $\Delta$ CFI = 0.000;  $\Delta$ RMSEA =  $-0.009$ ). These results indicate full scalar invariance across age groups (for more information regarding all models, see Table 8).

**Table 8.** Measurement Invariance of the Single-Factor Scales Across Age Groups.

Model	$\Delta$ df	$\Delta\chi^2$	$p$	$\Delta$ CFI	$\Delta$ RMSEA	Conclusion
Reduced 5-Item Version						
Configural	-	-	-	-	-	Supported
Metric	4	2.060	0.735	0.003	$-0.018$	Supported
Scalar	4	3.685	0.450	0.000	$-0.009$	Supported

Table 8. Cont.

Model	$\Delta df$	$\Delta\chi^2$	$p$	$\Delta CFI$	$\Delta RMSEA$	Conclusion
9-Item Version by Sum et al. [15]						
Configural	-	-	-	-	-	Supported
Metric	8	1.619	0.991	0.013	-0.010	Supported
Scalar	8	19.288	0.013	-0.007	-0.003	Supported *
Partial Scalar	7	13.100	0.070	-0.004	-0.003	Supported
15-Item Version						
Configural	-	-	-	-	-	Supported
Metric	14	11.086	0.679	0.001	-0.005	Supported
Scalar	14	23.952	0.046	-0.002	-0.002	Supported *
Partial Scalar	13	22.024	0.055	-0.002	-0.002	Supported

\* Scalar invariance across age may be essentially acceptable, indicated by the small  $\Delta CFI$  and  $\Delta RMSEA$  values. Due to the high number of participants, the  $\chi^2$ -Test was significant. For completion and comparison, we also report the results for the version freeing one item (Item 14 for the EFA Model and Item 13 for the Sum et al. Model) for partial scalar invariance.

#### 4. Discussion

In the present study, we aimed to examine the factorial validity, reliability, and longitudinal stability of a Spanish adaptation of the original 18-item Perceived Physical Literacy Instrument (PPLI) in adolescents aged 11–18 years. By retaining the full original item pool, we re-evaluated the dimensional structure of the instrument within a different cultural and developmental context and determined whether the multidimensional structure originally identified in physical education teachers [15] was adequate in a Spanish adolescent sample. Although exploratory analyses initially supported a three-factor solution broadly consistent with the conceptual domains of Whitehead's physical literacy (PL), some limitations were identified. Subsequent confirmatory analyses consistently indicated that a strong general factor accounted for the majority of shared variance among items, which may indicate a one-dimensional structure. Based on this finding, we developed a short 5-item version with acceptable validity and reliability for potential use in Spanish adolescents. However, due to the complexity of the PL concept, brief questionnaires might fail to measure the full philosophical breadth of the construct. Thus, the 5-item version must be understood as a brief screening-oriented measure of perceived PL, capturing a core general dimension but not the totality of the construct.

With respect to the three-dimensional structure, the Exploratory Factor Analysis (EFA) showed that 15 items provided the best fit in this sample of Spanish adolescents. This stands in contrast to the 9-item structure found by Sum et al. [15,16]. These differences may reflect contextual and educational factors rather than simple psychometric weakness. PL should be defined within specific social and pedagogical environments, and the meanings attached to movement, sport, health, autonomy, and social interaction may differ across countries and cultures. Similar to the findings of previous validation studies [15–17], our EFA results indicated three first-order factors. The first factor, which may be labelled Perceived Physical Competence and Self-Regulation, including items 1, 2, 3, 7, and 8, primarily assesses adolescents' perceptions of motor competence, physical fitness, and self-regulation of physical activity behavior. Conceptually, this domain overlaps substantially with the sense of self and self-confidence dimension described by Sum et al. [15,16], particularly with respect to confidence in movement and perceived competence. The second dimension may be labelled Adaptive and Social Competence, including items 6, 10, 11, 12, and 13. It integrates social and communication competence, environmental adaptability, and the functional application of knowledge. It closely resembles the self-expression and communication with others domain identified by Sum et al. [15,16], but extends it by incorporating elements of adaptability and long-term knowledge application that were previously associated with the knowledge and understanding dimension in Sum et al.'s [15,16] version. The third

dimension, which may be termed Motivation for Lifelong Engagement in Sport, consisted of items 9,14, 15, 16, and 18. This domain reflects intrinsic orientation toward healthy, long-term commitment to physical activity. While motivational aspects were embedded across the original three domains proposed by Sum et al. [15,16], this cluster represents a more explicit and cohesive motivational-behavioral orientation toward sustained participation.

Although these three domains partially overlap with the three-factor structure originally proposed by Sum et al. [15,16] (sense of self and self-confidence, self-expression and communication with others, and knowledge and understanding), the present findings suggest a more integrated configuration. In particular, the knowledge component appeared embedded in the competence and adaptive domains, instead of isolated as a proper domain. This pattern is consistent with Whitehead's conceptualization of physical literacy as a holistic and integrated concept [2]. However, factorial structures may vary depending on educational contexts and language adaptation processes, suggesting potential cultural influences of the Spanish culture and physical education context. In this regard, the current Spanish education law, which positions physical literacy as a core aim for promoting active and healthy lifestyles [4–6], may have influenced the present results. Future research in Spanish adolescent populations should further explore this hypothesis.

To compare the proposed 15-item version and Sum et al.'s [15,16] 9-item version, we analyzed their model fit with the present dataset. Overall, the 9-item model also exhibited adequate fit, which supports the notion that the same item may be interchangeably allocated into different dimensions and, consequently, that the multidimensionality of the scale may be controversial. This also supports the re-examination of the full content domain within the target population rather than assuming cross-contextual invariance of item functioning. In this regard, Spanish adolescents may differ substantially from physical education professionals from Hong Kong. Therefore, the psychometric reduction conducted in teachers may not fully generalize to youth populations. By retaining and re-evaluating all 18 items, this study provides new evidence that some previously discarded indicators may contribute meaningfully to the general construct when examined in adolescents. This reinforces the importance of population-specific validation rather than direct transplantation of shortened versions.

Although this three-factor structure showed acceptable fit, aligning with previous validations of the PPLI in both adult teachers and adolescents, subsequent confirmatory analyses consistently revealed some limitations that should be considered when using this structure. Specifically, very high loadings and high correlations between items and factors are commonly interpreted as evidence of essential unidimensionality rather than meaningful multidimensionality [47,50]. Furthermore, the internal consistency observed for the 15-item version ( $\omega = 0.919$ ) is relatively high. Although this indicates excellent reliability, it may also reflect some degree of item redundancy or content overlap, which should be considered in future applications. Therefore, while the three factor versions are valid, a unidimensional version of the PPLI may be adequate based on this study's data. However, due to the complexity and multidimensional nature of PL, psychometric essential unidimensionality at the instrument level should not be interpreted as theoretical reductionism at the construct level.

Given such dominance of the general factor at the measurement level and with the aim of providing a reduced instrument suitable for specific settings, a 5-item unidimensional short form was developed and cross-validated. The short form demonstrated high factor loadings ( $\lambda = 0.70$ – $0.79$ ), adequate convergent validity ( $AVE = 0.558$ ), high composite reliability ( $\omega = 0.821$ ), a similar test-retest reliability to the other tested models ( $ICC = 0.69$ ), and full longitudinal configural, metric, and scalar invariance. Importantly, the five-item version achieved temporal stability comparable to the 9- and 15-item versions, despite

containing substantially fewer items. This indicates that the short form preserves essential psychometric properties while substantially reducing respondent burden, which may be useful for large-scale epidemiological studies, school-based screening, or intervention monitoring where assessment time is constrained.

The content of this 5-item version mainly focuses on the domains of perceived physical competence, self-regulation, and motivation for lifelong physical activity. In contrast, items representing social and adaptive competence in physical activity contexts were not retained in the short form. In addition, compared with the three domains originally proposed by Sum et al. [15,16], the short form strongly reflects sense of self and self-confidence, as well as knowledge and understanding. However, again, it does not reflect self-expression and communication with others. This pattern suggests that the short form is assessing the core components of perceived physical literacy, with emphasis on perceived physical competence, self-regulatory capacity, and motivational orientation toward lifelong engagement in physical activity, but excluding the social and communicative skill assessment.

According to our findings, the 9- and 15-item versions may be valid and reliable, so researchers and educators may select the instrument version according to their objectives. The 15-item three-factor model may be useful when interest lies in exploring domain-specific patterns, while the five-item unidimensional short form may be recommended when there is time restriction and the goal is to rapidly assess the global construct. The 9-item version can also be used to compare with other international studies that have used that structure. Another main difference among the three versions is the inclusion of the social component, which has been highlighted by previous studies as part of a broader, holistic view of the concept of PL [53,54]. However, empirical assessment tools often prioritize physical competence and motivation/confidence, with the social component appearing less consistently as an independent latent domain [14]. Therefore, when the aim is the assessment of PL with special emphasis on the social and communicative components, the 15-item version would be preferable because it includes three related items (Items 10, 11, and 16), whereas the 9-item version includes only one (Item 11) and the 5-item version includes none.

From a practical perspective, the promotion of PL in children and adolescents is a promising strategy for promoting lifelong engagement in physical activity and improving health outcomes. A wide range of movement experiences, including dance, fitness activities, games, gymnastics, individual sports, and outdoor activities, may contribute to the development of PL and, consequently, enhance physical activity participation and health-related outcomes [55]. In addition, interventions that extend beyond the school setting, incorporating family or home-based components, may provide further benefits [56]. Although the integration of PL into school curricula is encouraging, the effectiveness of these initiatives requires further empirical evaluation [55]. In this context, the PPLI may serve as a brief and efficient screening tool to assess PL before and after an intervention. The 5-item version, due to its short administration time, can be easily implemented at the beginning and end of a teaching unit to monitor changes in students' perceived physical literacy, particularly in terms of motivation, perceived competence, and engagement in physical activity. When a more comprehensive assessment is required, the 15-item version may be used to capture a broader range of domains.

Several limitations should be considered. First, item selection was guided by both statistical and theoretical considerations, so replication in independent samples may be recommended. Second, although the short form is valid and reliable, some PL domains are not assessed using this version, such as the social and communicative one. Furthermore, it does not permit the assessment of any possible subscale-specific variance, so respective interpretations are limited. Third, although the sample size was adequate and involved

public and private institutions from urban and rural areas of the western, northern, eastern and southern regions of Spain, representativeness may not be ensured. Despite these limitations, the current study provides an in-depth analysis of the PPLI structure, which is one of the PL assessment tools with the highest level of validity [14], based on the 18 initial items developed by Sum et al. [15] and comparing between three different options.

## 5. Conclusions

The 15- and 5-item versions were developed and cross-validated from the original 18-item version of the PPLI developed by Sum et al. [15]. They demonstrated adequate validity, reliability, and temporal stability. In addition, the 9-item version proposed by Sum et al. [15], and previously validated in Spanish [17], may also be valid and reliable, but the same limitations were found. Each version has its strengths and limitations: the 15-item version is the longest but allows domain-level interpretation, the 9-item version facilitates comparability with previous international research, and the 5-item version may be particularly useful in contexts with time constraints. However, this reduced version does not explicitly assess the social and communicative component of physical literacy, so it may not be the preferred choice for comprehensive assessment of physical literacy in clinical or detailed pedagogical diagnostic settings. Therefore, the choice of version depends on the specific context and objectives.

**Author Contributions:** Conceptualization, J.A.R.-M., J.C.A. and D.C.-M.; methodology, J.A.R.-M., J.F.-S., R.P.-C., M.M.-M., J.C.-V. and R.B.; formal analysis, E.F.-S. and R.B.; investigation, J.A.R.-M., R.P.-C., M.M.-M., I.G.-G. and J.C.-V.; resources, J.C.A., D.C.-M. and I.G.-G.; data curation, J.C.A., R.B. and E.F.-S.; writing—original draft preparation, J.A.R.-M., R.B., D.C.-M. and J.F.-S.; writing—review and editing, R.B., D.C.-M., I.G.-G., J.A.R.-M., J.F.-S., R.P.-C., M.M.-M. and J.C.-V.; supervision, D.C.-M., R.B., M.M.-M., J.C.-V. and J.C.A.; funding acquisition, D.C.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Dirección General de Ordenación del Juego, SUVB24/00031 (Spanish Ministry of Consumer Affairs). Author E.F.S. was hired as research staff under SUVB24/00031 Project. The author R.P.-C. was supported by a grant from the Spanish Ministry of Universities (FPU22/00262). The author M.M.-M. was supported by a grant from the Andalusian Regional Government/CUII and the ESF+ (DGP\_POST\_2024\_00985).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the University of Extremadura (Protocol code: 176//2025, and date of approval: 12 May 2025).

**Informed Consent Statement:** Written informed consent was obtained from all participants. In addition, informed consent was provided by the parents or legal guardians of all adolescent participants.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The dataset is not publicly available due to the privacy of the respondents.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AVE	Average Variance Extracted
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
EFA	Exploratory Factor Analysis
ICC	Intraclass Correlation Coefficient
KMO	Kaiser–Meyer–Olkin

LOMLOE	Ley Orgánica de Modificación de la Ley Orgánica de Educación
MLR	Robust Maximum Likelihood Estimation
PE	Physical Education
PL	Physical Literacy
PPLI	Perceived Physical Literacy Instrument
PPLI-Sp	Perceived Physical Literacy Instrument—Spanish Version
QR	Quick Response
RMSEA	Root Mean Square Error of Approximation
SD	Standard Deviation
SEM	Standard Error of Measurement
SRD	Smallest Real Difference
SRMR	Standardized Root Mean Square Residual
TLI	Tucker–Lewis Index
WLSMV	Weighted Least Squares Mean and Variance Adjusted

## References

1. Edwards, L.C.; Bryant, A.S.; Keegan, R.J.; Morgan, K.; Jones, A.M. Definitions, Foundations and Associations of Physical Literacy: A Systematic Review. *Sports Med.* **2016**, *47*, 113–126. [[CrossRef](#)]
2. Whitehead, M. *Physical Literacy: Throughout the Lifecourse*; Routledge: Oxfordshire, UK, 2010; pp. 1–230. [[CrossRef](#)]
3. Barratt, J.; Dudley, D.; Stylianou, M.; Cairney, J. A Conceptual Model of an Effective Early Childhood Physical Literacy Pedagogue. *J. Early Child. Res.* **2024**, *22*, 381–394. [[CrossRef](#)]
4. World Health Organization. *Global Action Plan on Physical Activity 2018–2030: More Active People for a Healthier World*; WHO: Geneva, Switzerland, 2018.
5. Jefatura del Estado (España). Ley Orgánica 3/2020, de 29 de Diciembre, Por La Que Se Modifica La Ley Orgánica 2/2006, de 3 de Mayo, de Educación. *Boletín Oficial del Estado (BOE-A-2020-17264)*, 30 December 2020.
6. McLennan, N. *Making the Case for Inclusive Quality Physical Education Policy Development: A Policy Brief*; UNESCO Publishing: Paris, France, 2021.
7. Reilly, J.J.; Barnes, J.; Gonzalez, S.; Huang, W.Y.; Manyanga, T.; Tanaka, C.; Tremblay, M.S. Recent Secular Trends in Child and Adolescent Physical Activity and Sedentary Behavior Internationally: Analyses of Active Healthy Kids Global Alliance Global Matrices 1.0 to 4.0. *J. Phys. Act. Health* **2022**, *19*, 729–736. [[CrossRef](#)] [[PubMed](#)]
8. Guthold, R.; Stevens, G.A.; Riley, L.M.; Bull, F.C. Global Trends in Insufficient Physical Activity among Adolescents: A Pooled Analysis of 298 Population-Based Surveys with 1.6 Million Participants. *Lancet Child Adolesc. Health* **2020**, *4*, 23–35. [[CrossRef](#)] [[PubMed](#)]
9. Brown, D.M.Y.; Dudley, D.A.; Cairney, J. Physical Literacy Profiles Are Associated with Differences in Children’s Physical Activity Participation: A Latent Profile Analysis Approach. *J. Sci. Med. Sport* **2020**, *23*, 1062–1067. [[CrossRef](#)] [[PubMed](#)]
10. Jiang, T.; Zhao, G.; Fu, J.; Sun, S.; Chen, R.; Chen, D.; Hu, X.; Li, Y.; Shen, F.; Hong, J.; et al. Relationship Between Physical Literacy and Cardiorespiratory Fitness in Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Med.* **2024**, *55*, 473–485. [[CrossRef](#)]
11. Barratt, J.; Goss, H.; Erskine, N.; James, M.; Töpfer, C.; Pfeifer, K.; Cairney, J.; Carl, J. Experiences, Influencing Factors, and Perceived Outcomes from Physical Literacy Interventions: A Qualitative Meta-Synthesis. *Int. J. Qual. Stud. Health Well-Being* **2026**, *21*, 2613973. [[CrossRef](#)]
12. Vanhelst, J.; Béghin, L.; Drumez, E.; Labreuche, J.; Polito, A.; De Ruyter, T.; Censi, L.; Ferrari, M.; Miguel-Berges, M.L.; Michels, N.; et al. Changes in Physical Activity Patterns from Adolescence to Young Adulthood: The BELINDA Study. *Eur. J. Pediatr.* **2023**, *182*, 2891. [[CrossRef](#)]
13. Cairney, J.; Dudley, D.; Kwan, M.; Bulten, R.; Kriellaars, D. Physical Literacy, Physical Activity and Health: Toward an Evidence-Informed Conceptual Model. *Sports Med.* **2019**, *49*, 371–383. [[CrossRef](#)]
14. Jean de Dieu, H.; Zhou, K. Physical Literacy Assessment Tools: A Systematic Literature Review for Why, What, Who, and How. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7954. [[CrossRef](#)]
15. 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. Construction and Validation of a Perceived Physical Literacy Instrument for Physical Education Teachers. *PLoS ONE* **2016**, *11*, e0155610. [[CrossRef](#)]
16. Sum, K.W.R.; Cheng, C.F.; Wallhead, T.; Kuo, C.C.; Wang, F.J.; Choi, S.M. Perceived Physical Literacy Instrument for Adolescents: A Further Validation of PPLI. *J. Exerc. Sci. Fit.* **2018**, *16*, 26–31. [[CrossRef](#)] [[PubMed](#)]

17. Mendoza-Muñoz, M.; Carlos-Vivas, J.; Castillo-Paredes, A.; Sum, R.K.W.; Rojo-Ramos, J.; Pastor-Cisneros, R. Translation, Cultural Adaptation and Validation of Perceived Physical Literacy Instrument-Spanish Version (PPLI-Sp) for Adults. *J. Sports Sci. Med.* **2023**, *22*, 455–464. [CrossRef]
18. López-Gil, J.F.; Martínez-Vizcaíno, V.; Tárraga-López, P.J.; García-Hermoso, A. Cross-Cultural Adaptation, Reliability, and Validation of the Spanish Perceived Physical Literacy Instrument for Adolescents (S-PPLI). *J. Exerc. Sci. Fit.* **2023**, *21*, 246–252. [CrossRef]
19. Whitehead, M. *Physical Literacy Across the World*; Taylor & Francis: London, UK, 2019; pp. 1–290. [CrossRef]
20. Resneck, J.S. Revisions to the Declaration of Helsinki on Its 60th Anniversary: A Modernized Set of Ethical Principles to Promote and Ensure Respect for Participants in a Rapidly Innovating Medical Research Ecosystem. *JAMA* **2024**, *333*, 15–17. [CrossRef]
21. Beaton, D.E.; Bombardier, C.; Guillemin, F.; Ferraz, M.B. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine* **2000**, *25*, 3186–3191. [CrossRef] [PubMed]
22. Harkness, J.A.; Villar, A.; Edwards, B. Translation, Adaptation, and Design. In *Survey Methods in Multicultural, Multinational, and Multiregional Contexts*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; pp. 115–140. [CrossRef]
23. Bentler, P.M.; Bonett, D.G. Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychol. Bull.* **1980**, *88*, 588–606. [CrossRef]
24. Hambleton, R.K.; Merenda, P.F.; Spielberger, C.D. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*; Psychology Press: London, UK, 2004.
25. Behr, D. Assessing the Use of Back Translation: The Shortcomings of Back Translation as a Quality Testing Method. *Int. J. Soc. Res. Methodol.* **2017**, *20*, 573–584. [CrossRef]
26. Worthington, R.L.; Whittaker, T.A. Scale Development Research. *Couns. Psychol.* **2006**, *34*, 806–838. [CrossRef]
27. Floyd, F.J.; Widaman, K.F. Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychol. Assess.* **1995**, *7*, 286–299. [CrossRef]
28. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*, 2nd ed.; The Guilford Press: New York, NY, USA, 2015.
29. R Core Team. *R A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025. Available online: <https://www.R-project.org/> (accessed on 15 January 2026).
30. Posit Team. *RStudio: Integrated Development Environment for R. Posit Software*; PBC: Boston, MA, USA, 2024. Available online: <http://www.posit.co/> (accessed on 15 January 2026).
31. Rosseel, Y. Lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [CrossRef]
32. Rosseel, Y.; Jorgensen, T.D.; De Wilde, L. *Latent Variable Analysis, R Package Lavaan Version 0.6-21*; CRAN: Contributed Packages; CRAN: Vienna, Austria, 2025. [CrossRef]
33. Jorgensen, T.D.; Pornprasertmanit, S.; Schoemann, A.M.; Rosseel, Y. *SemTools: Useful Tools for Structural Equation Modeling (Version 0.5-7)* [R Package]. 2025. Available online: <https://CRAN.R-project.org/package=semTools> (accessed on 15 January 2026).
34. Bollen, K.A. *Structural Equations with Latent Variables*; Wiley-Interscience: New York, NY, USA, 2014; pp. 1–514. [CrossRef]
35. Hu, L.T.; Bentler, P.M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Struct. Equ. Model.* **1999**, *6*, 1–55. [CrossRef]
36. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 4th ed.; Guilford Press: New York, NY, USA, 2015.
37. Costello, A.B.; Osborne, J.W. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Pract. Assess. Res. Eval.* **2005**, *10*, 7. [CrossRef]
38. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 8th ed.; Cengage Learning: Andover, UK, 2019.
39. Fabrigar, L.R.; MacCallum, R.C.; Wegener, D.T.; Strahan, E.J. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychol. Methods* **1999**, *4*, 272–299. [CrossRef]
40. Browne, M.W.; Cudeck, R. Alternative Ways of Assessing Model Fit. *Sociol. Methods Res.* **1992**, *21*, 230–258. [CrossRef]
41. Xia, Y.; Yang, Y. RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods. *Behav. Res. Methods* **2018**, *51*, 409–428. [CrossRef]
42. Fornell, C.; Larcker, D.F. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Mark. Res.* **1981**, *18*, 39–50. [CrossRef]
43. Flora, D.B.; Curran, P.J. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data. *Psychol. Methods* **2004**, *9*, 466–491. [CrossRef]
44. Dunn, T.J.; Baguley, T.; Brunsden, V. From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation. *Br. J. Psychol.* **2014**, *105*, 399–412. [CrossRef] [PubMed]
45. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]
46. Aldridge, V.K.; Dovey, T.M.; Wade, A. Assessing Test-Retest Reliability of Psychological Measures. *Eur. Psychol.* **2017**, *22*, 207–218. [CrossRef]

47. Reise, S.P. The Rediscovery of Bifactor Measurement Models. *Multivar. Behav. Res.* **2012**, *47*, 667–696. [[CrossRef](#)]
48. Sass, D.A.; Schmitt, T.A.; Marsh, H.W. Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Struct. Equ. Model.* **2014**, *21*, 167–180. [[CrossRef](#)]
49. Blum, R.W.; Astone, N.M.; Decker, M.R.; Mouli, V.C. A Conceptual Framework for Early Adolescence: A Platform for Research. *Int. J. Adolesc. Med. Health* **2014**, *26*, 321–331. [[CrossRef](#)] [[PubMed](#)]
50. Bonifay, W.; Lane, S.P.; Reise, S.P. Three Concerns with Applying a Bifactor Model as a Structure of Psychopathology. *Clin. Psychol. Sci.* **2017**, *5*, 184–186. [[CrossRef](#)]
51. Thomas, S.; DeVellis, R.F.; Thorpe, C.T. Scale Development: Theory and Applications. *Pers. Psychol.* **2022**, *75*, 243–244. [[CrossRef](#)]
52. Kenny, D.A.; Kaniskan, B.; McCoach, D.B. The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociol. Methods Res.* **2015**, *44*, 486–507. [[CrossRef](#)]
53. 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.; et al. Defining Physical Literacy for Application in Australia: A Modified Delphi Method. *J. Teach. Phys. Educ.* **2019**, *38*, 105–118. [[CrossRef](#)]
54. Fortnum, K.; Weber, M.D.; Dudley, D.; Tudella, E.; Kwan, M.; Richard, V.; Cairney, J. Physical Literacy, Physical Activity, and Health: A Citation Content Analysis and Narrative Review. *Sports Med.-Open* **2025**, *11*, 44. [[CrossRef](#)]
55. Grauduszus, M.; Koch, L.; Wessely, S.; Joisten, C. School-Based Promotion of Physical Literacy: A Scoping Review. *Front. Public Health* **2024**, *12*, 1322075. [[CrossRef](#)]
56. Nezondet, C.; Gandrieau, J.; Bourrelie, J.; Nguyen, P.; Zunquin, G. The Effectiveness of a Physical Literacy-Based Intervention for Increasing Physical Activity Levels and Improving Health Indicators in Overweight and Obese Adolescents (CAPACITES 64). *Children* **2023**, *10*, 956. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.