Contents list available at PELTA

Physical Education Learning, Teaching, and Assessment (PELTA)

Publisher: Perkumpulan Pengelola Jurnal Keolahragaan Indonesia (PJKI)

Preliminary psychometric evidence for PLAYfun and PLAYself among Indonesian elementary school students: Item validity and internal consistency

Bukti psikometrik awal PLAYfun dan PLAYself pada siswa sekolah dasar Indonesia: Validitas butir dan konsistensi internal

Ropik Hidayat[1,*], Heny Setyawati[2], Siti Baitul Mukkaromah[2]

[1] *SMA Negeri 1 Kesesi, Indonesia*
[2] *Universitas Negeri Semarang, Indonesia*
email: ropik1493@gmail.com

ARTICLE INFO

Received 19 July 2026
Revised 29 August 2026
Accepted 15 September 2026
Published 30 September 2026

Keywords:
Internal consistency
Item validity
Physical literacy
PLAYfun
PLAYself

***Kata kunci*:**
Konsistensi internal
Literasi fisik
PLAYfun
PLAYself
Validitas butir

APA style in citing this article:
Hidayat, R., Setyawati, H., & Mukarromah, S. B. (2026). Preliminary psychometric evidence for PLAYfun and PLAYself among Indonesian elementary school students: Item validity and internal consistency. *Physical Education Learning, Teaching, and Assessment (PELTA), 1*(1), 8-14. https://doi.org/10.68536/pelta.v1i1.17

ABSTRACT

Background: Physical literacy encompasses motivation, confidence, physical competence, knowledge, and understanding that support lifelong engagement in physical activity. Assessing physical literacy in children requires instruments that are contextually relevant and demonstrate adequate psychometric properties.
Objective: This study aimed to provide preliminary psychometric evidence for culturally adapted versions of PLAYfun and PLAYself in the Indonesian context by examining item–total relationships and internal consistency among elementary school students.
Methods: A total of 71 fifth-grade students aged 10–11 years from three elementary schools in Pekalongan Regency participated in the study. PLAYfun comprised 18 motor-skill assessment items, whereas PLAYself consisted of 21 self-report items. Two PLAYself activities were modified to improve their geographical and cultural relevance to the Indonesian context. Item–total relationships were examined using Pearson correlations and corrected item–total correlations (CITC), while internal consistency was assessed using Cronbach's alpha.
Results: For PLAYfun, item–total correlations ranged from 0.504 to 0.893, with very high internal consistency (α = 0.955). For PLAYself, item–total correlations ranged from 0.265 to 0.912, with high internal consistency (α = 0.895). However, five PLAYself items (PS5, PS6, PS7, PS8, and PS19) showed CITC values below 0.30, indicating relatively weak contributions to scale consistency and suggesting the need for further evaluation.
Conclusion: These findings provide preliminary psychometric support for PLAYfun and PLAYself within the elementary-school context examined; however, further validation involving larger and more diverse samples, factor-structure analyses, and additional reliability testing is required before broader use.

ABSTRAK

Latar belakang: *Physical literacy* mencakup motivasi, kepercayaan diri, kompetensi fisik, pengetahuan, dan pemahaman yang mendukung keterlibatan dalam aktivitas fisik sepanjang hayat. Pengukuran physical literacy pada anak memerlukan instrumen yang tidak hanya relevan secara kontekstual, tetapi juga menunjukkan karakteristik psikometrik yang memadai.
Tujuan: Penelitian ini bertujuan untuk memperoleh bukti psikometrik awal PLAYfun dan PLAYself yang telah diadaptasi untuk konteks Indonesia, dengan mengevaluasi korelasi item-total dan konsistensi internal pada siswa sekolah dasar.
Metode: Sebanyak 71 siswa kelas V berusia 10–11 tahun dari tiga sekolah dasar di Kabupaten Pekalongan berpartisipasi dalam penelitian ini. PLAYfun terdiri atas 18 item penilaian keterampilan motorik, sedangkan PLAYself terdiri atas 21 item laporan diri. Dua aktivitas pada PLAYself dimodifikasi agar lebih relevan dengan konteks geografis dan budaya Indonesia. Hubungan setiap item dengan skor total dianalisis menggunakan korelasi *Pearson* dan *corrected item–total correlation* (CITC), sedangkan konsistensi internal dievaluasi menggunakan *Cronbach's alpha*.

Hasil: Pada PLAYfun, korelasi item-total berkisar antara 0,504–0,893 dan konsistensi internal sangat tinggi (α = 0,955). Pada PLAYself, korelasi item-total berkisar antara 0,265–0,912 dan konsistensi internal juga tinggi (α = 0,895). Namun, lima item PLAYself (PS5, PS6, PS7, PS8, dan PS19) memiliki CITC < 0,30, yang menunjukkan kontribusi relatif lemah terhadap konsistensi skala dan memerlukan evaluasi lebih lanjut.

Kesimpulan: Temuan ini memberikan bukti psikometrik awal yang mendukung penggunaan PLAYfun dan PLAYself pada konteks sekolah dasar yang diteliti, tetapi validasi lebih lanjut menggunakan sampel yang lebih besar dan beragam, analisis struktur faktor, serta pengujian reliabilitas tambahan diperlukan sebelum penggunaan yang lebih luas.

© 2026 The Author. Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Highlights

- ✓ PLAYfun demonstrated very high internal consistency among Indonesian elementary school students.
- ✓ PLAYself showed high internal consistency, although five items had weak corrected item–total correlations.
- ✓ Contextual adaptation of PLAYself provides preliminary support for its use in the Indonesian elementary-school setting.

Introduction

Physical literacy is a multidimensional concept encompassing motivation, confidence, physical competence, knowledge, and understanding that support lifelong engagement in physical activity. Among elementary school children, physical literacy is particularly important because this period represents a critical stage for motor skill development, the formation of perceived competence, and the establishment of physical activity habits that may persist across the lifespan. Cairney et al. (2019) conceptualized physical literacy as a construct associated with physical activity and health, while Caldwell et al. (2020) demonstrated associations between physical literacy, physical activity, and health indicators in school-aged children. Accordingly, the development of physical literacy is not only relevant to physical education but also contributes to establishing the foundations of an active lifestyle from an early age.

A further concern relates to how physical literacy can be assessed accurately. Assessment instruments need to capture more than motor competence alone, as physical literacy also includes affective and cognitive dimensions. In a systematic review of physical literacy assessment tools for school-aged children, Barnett et al. (2023) reported substantial variability in validity, reliability, and feasibility across available instruments, emphasizing the importance of considering both psychometric evidence and population context when selecting an assessment tool. Similarly, Shearer et al. (2021a) highlighted that existing physical literacy instruments differ in the extent to which they assess physical, affective, and cognitive domains. These findings suggest that instruments developed and validated in one country cannot necessarily be applied directly to another population without evaluating their psychometric and contextual suitability.

One widely used assessment framework is the Physical Literacy Assessment for Youth (PLAY), particularly PLAYfun and PLAYself. PLAYfun is designed to assess movement competence through direct observation of motor skills, whereas PLAYself is a self-report instrument that captures children's perceptions of their own physical literacy. Previous evidence has demonstrated promising psychometric performance. Stearns et al. (2019) reported moderate-to-good inter-rater reliability for PLAYfun, with intraclass correlation coefficients (ICCs) of 0.78 and 0.82 across two assessment occasions. Meanwhile, Jefferies et al. (2021) found that PLAYself demonstrated psychometric properties supporting its use as a self-reported measure of physical literacy, including high test–retest reliability and satisfactory fit to the Rasch model. Collectively, these findings indicate that PLAYfun and PLAYself have a relatively strong empirical foundation in the populations in which they were originally examined.

Other studies have also emphasized that physical literacy assessment should account for population and cultural context. Gunnell et al. (2018) refined the Canadian Assessment of Physical Literacy using theory-driven approaches and factor analyses, while Knisel et al. (2024) validated the CAPL-2 questionnaire among schoolchildren in Germany and Poland. Scott et al. (2021) further demonstrated that physical literacy frameworks may be shaped by national education and sport policy contexts. Taken together, these studies indicate that physical literacy instruments require not only a sound theoretical basis but also careful evaluation when applied across different linguistic, cultural, environmental, and physical activity contexts.

Despite these developments, several research gaps remain. First, psychometric evidence for PLAYfun and PLAYself has been generated predominantly in international settings, whereas evidence regarding their use among Indonesian elementary school students remains limited. Second, some activities included in PLAYself reflect the environmental and cultural context in which the instrument was originally developed and may not correspond closely to the lived experiences of Indonesian children. In the present study, for example, activities such as ice skating and cross-country skiing were considered less relevant to the geographical context and physical activity experiences of Indonesian students and therefore required contextual modification. Third, although previous studies have reported evidence regarding the reliability and validity of PLAY instruments in their original populations, preliminary evidence concerning item–total relationships and internal consistency following contextual adaptation among Indonesian elementary school students remains insufficient. Therefore, it is necessary to examine whether the internal characteristics of PLAYfun and PLAYself remain adequate when these instruments are applied in the Indonesian context.

The novelty of the present study lies in providing preliminary psychometric evidence for PLAYfun and PLAYself among Indonesian elementary school students following contextual adaptation of selected PLAYself activities to better reflect local geographical and cultural conditions. This study was not intended to provide national validation or a comprehensive assessment of construct validity. Rather, it represents an initial stage of evaluation focusing on item–total relationships and internal consistency within the population examined. Specifically, two PLAYself activities were modified: ice skating was replaced with swimming in rivers or the sea, while cross-country skiing was replaced with cycling on rural or mountainous roads. These modifications were developed

through discussions with physical education experts, physical education teachers, and physical literacy researchers.

Based on these research gaps, this study aimed to evaluate preliminary psychometric evidence for PLAYfun and PLAYself among Indonesian elementary school students by examining item–total relationships and internal consistency, while also describing the contextual adaptation of selected PLAYself items. Theoretically, this study is expected to extend current evidence regarding the use of PLAY instruments across different cultural contexts. Methodologically, the findings may provide a basis for subsequent investigations involving factor analysis, test–retest reliability, and inter-rater reliability. From a practical perspective, the results may provide preliminary information for physical education teachers and researchers considering the use of PLAYfun and PLAYself to assess physical literacy among Indonesian elementary school students.

Materials and Methods

Study Design

This study employed a quantitative approach using a preliminary psychometric study design to evaluate the internal characteristics of the Physical Literacy Assessment for Youth (PLAY), comprising PLAYfun and PLAYself, when administered to Indonesian elementary school students. The evaluation focused on three aspects: contextual adaptation of the instruments, relationships between individual item scores and total scores, and internal consistency. Accordingly, the present study was not intended to provide a comprehensive assessment of construct validity, as it did not include confirmatory factor analysis, external criterion validity testing, or test–retest reliability assessment. The analyses were limited to item–total correlations and Cronbach's alpha; therefore, the terms "construct validity" and "criterion validity" were not used in the final interpretation of the findings.

Data were collected during the second semester of the 2024/2025 academic year at three elementary schools in Pekalongan Regency, Indonesia. The schools were purposively selected based on their willingness to participate and the suitability of their student populations for the objectives of the study.

Participants

The study involved 71 fifth-grade students aged 10–11 years from three elementary schools in Pekalongan Regency. Fifth-grade students were selected because they were considered to be at a relatively comparable stage of motor development and to have had sufficient prior exposure to school-based physical education.

Interventions

PLAYfun is an observational assessment comprising 18 motor-skill items covering locomotor skills, upper- and lower-body object-control skills, and balance/stability (Longmuir et al., 2018). PLAYself is a 21-item self-report questionnaire designed to assess students' perceived physical literacy, including dimensions of motivation, confidence, physical competence, and understanding (Stearns et al., 2019).

The PLAY assessment framework classifies performance into four developmental stages: Initial (0–25), Emerging (26–50), Competent (51–75), and Proficient (76–100) (Kriellaars & Robillard, 2013). PLAYfun requires approximately 15–20 minutes to administer and can be conducted with one to three students at a time.

PLAYself was scored in accordance with the procedures specified for the instrument. Responses to the environmental activity items (Q1–Q6) were scored as follows: never tried = 0, not very good = 25, fairly good = 50, very good = 75, and extremely good = 100. For the physical literacy self-description items (Q7–Q19, except Q13), responses were scored as follows: not at all true = 0, usually not true = 33, true = 67, and very true = 100. Item 13, "I worry about trying a new sport or activity," was reverse-scored: not at all true = 100, usually not true = 67, true = 33, and very true = 0. For the relative physical literacy ratings (Q19–Q21), responses were scored as follows: strongly disagree = 0, disagree = 33, agree = 67, and strongly agree = 100 (Kriellaars & Robillard, 2013). Data from the fitness subsection were not included because no scoring procedure for this subsection was specified in the PLAYself protocol.

PLAYfun was administered by assessors who had received training in accordance with the standardized assessment protocol. PLAYself was completed by the students under the supervision of the researchers and physical education teachers to ensure that each item was properly understood.

Statistical analysis

Pearson product–moment correlations were used to examine the relationship between each item score and the corresponding total score. The calculated correlation coefficients were compared with the critical value at the 5% significance level ($r = 0.232$; $N = 71$). An item was considered to demonstrate an adequate item–total relationship when its correlation coefficient exceeded the critical value.

Internal consistency was assessed using Cronbach's alpha. An instrument was considered reliable when the alpha coefficient exceeded 0.60 (Sugiyono, 2017). The interpretation of reliability coefficients followed Arikunto (2010): 0.00–0.20 = very low, 0.20–0.40 = low, 0.40–0.60 = moderate, 0.60–0.80 = high, and 0.80–1.00 = very high.

The use of both instruments was further supported by evidence from previous studies. PLAYfun has demonstrated good inter-rater reliability (ICC = 0.78–0.82) and high internal consistency ($\alpha = 0.83–0.87$) (Stearns et al., 2019). PLAYself has also demonstrated satisfactory fit to the Rasch model, high test–retest reliability (ICC = 0.81–0.84), and internal consistency of $\alpha = 0.80$ (Jefferies et al., 2021).

All statistical analyses were performed using IBM SPSS Statistics version 26.0 for Windows. Corrected item–total correlation (CITC) was used to assess the contribution of each item to the overall scale. Items with CITC values below 0.30 were identified as candidates for further evaluation or revision in subsequent stages of instrument development.

Results

The distribution of physical literacy levels based on PLAYfun showed that most students were classified in the Emerging category, comprising 55 of 71 participants (77.5%). Nine students (12.7%) were classified in the Initial category, whereas only seven students (9.8%) were classified as Competent. No participants were classified in the Proficient category.

As shown in Table 1, the majority of participants were classified at the Emerging stage, indicating that PLAYfun scores in this sample were concentrated at an intermediate developmental level. The proportion of students who reached the Competent category was relatively small, and none were classified as Proficient. These findings are descriptive and characterize the profile of the study sample rather than provide evidence regarding the validity of the instrument.

Table 2 shows that the mean scores of the individual PLAYfun items varied across items. PF8 had the highest mean

Hidayat, Setyawati, & Mukarromah; PELTA 1(1) 2026 8-14 ISSN 3164-7987 (Online)

score (42.61), whereas PF17 had the lowest (34.37). Variability also differed across items, with the largest standard deviation observed for PF8 (SD = 14.925) and the smallest for PF18 (SD = 10.681). These descriptive statistics provide an overview of the distribution of responses across items but do not, by themselves, indicate item-level psychometric quality, which was further evaluated using item–total correlations and corrected item–total correlations (CITC).

As presented in Table 3, all PLAYfun items showed CITC values above 0.30, indicating that none of the items demonstrated a weak corrected relationship with the total score. PF12 had the highest CITC value (0.879), whereas PF3 had the lowest (0.504); however, both remained above the evaluative threshold used in this study. Overall, Cronbach's α = 0.955 indicated very high internal consistency. Nevertheless, an alpha coefficient approaching 1.00 should be interpreted cautiously, as it may also suggest substantial similarity or possible redundancy among some items.

Table 1
Distribution of PLAYfun score categories

Category	Score range	n	%
Initial	0–25	9	12.7
Emerging	26–50	55	77.5
Competent	51–75	7	9.8
Proficient	76–100	0	0.0
Total		71	100.0

Source: Study data (2024)

Table 2
Descriptive statistics for PLAYfun items

Item	Mean	SD	n	Item	Mean	SD	n
PF1	37.25	11.674	71	PF10	34.52	12.584	71
PF2	42.18	13.542	71	PF11	38.24	12.624	71
PF3	41.27	12.893	71	PF12	36.41	12.961	71
PF4	39.08	11.811	71	PF13	36.27	12.669	71
PF5	37.70	13.813	71	PF14	35.70	13.344	71
PF6	37.23	13.012	71	PF15	35.85	12.705	71
PF7	35.35	12.910	71	PF16	35.99	13.132	71
PF8	42.61	14.925	71	PF17	34.37	13.173	71
PF9	39.93	12.719	71	PF18	41.27	10.681	71

Source: IBM SPSS Statistics version 26.0

Table 3
Item–total correlations and corrected item–total correlations for PLAYfun

Item	Pearson item-total *r*	CITC	CITC interpretation
PF1	0.621	0.570	Adequate
PF2	0.808	0.768	Strong
PF3	0.570	0.504	Adequate
PF4	0.871	0.846	Very strong
PF5	0.847	0.810	Very strong
PF6	0.719	0.671	Strong
PF7	0.629	0.577	Adequate
PF8	0.783	0.741	Strong
PF9	0.775	0.741	Strong
PF10	0.677	0.647	Strong
PF11	0.837	0.812	Very strong
PF12	0.893	0.879	Very strong
PF13	0.727	0.702	Strong
PF14	0.750	0.712	Strong
PF15	0.815	0.797	Strong
PF16	0.823	0.806	Very strong
PF17	0.640	0.610	Strong
PF18	0.783	0.783	Strong

Source: IBM SPSS Statistics version 26.0

As shown in Table 4, PLAYself demonstrated high overall internal consistency, with Cronbach's α = 0.895. Although all Pearson item–total correlation coefficients exceeded the critical value used in the study, five items—PS5, PS6, PS7, PS8, and PS19—had CITC values below 0.30. These values indicate that the five items contributed less strongly to overall scale consistency than the remaining items. Therefore, these items should not be considered to demonstrate psychometric performance equivalent to that of the other items and should instead be retained for further evaluation using larger samples, factor-structure analyses, and additional assessment of cultural relevance and item comprehension.

As summarized in Table 5, both instruments demonstrated high internal consistency within the present sample. PLAYfun showed a more consistent item pattern, as all CITC values exceeded 0.30. PLAYself also demonstrated a high alpha coefficient, although five items showed low CITC values and therefore warrant

further evaluation. Overall, the present findings provide stronger support for the internal consistency of PLAYfun, whereas PLAYself demonstrated promising results but requires further examination of several individual items.

Discussion

This study yielded two main findings. First, PLAYfun demonstrated very high internal consistency, with a Cronbach's alpha of 0.955, and all items showed corrected item–total correlation (CITC) values above 0.30. Second, PLAYself also demonstrated high internal consistency, with a Cronbach's alpha of 0.895; however, five items—PS5, PS6, PS7, PS8, and PS19—had CITC values below 0.30. Thus, PLAYfun showed a more consistent pattern across items, whereas PLAYself included several items that require further evaluation. These findings should be interpreted as preliminary psychometric evidence concerning item–total relationships and internal consistency rather than as comprehensive evidence of construct validity.

The findings for PLAYfun are consistent with those of Stearns et al. (2019), who reported good reliability characteristics for PLAYfun among children and adolescents, including adequate inter-rater reliability. These findings support the view that the item structure of PLAYfun is relatively consistent in representing observable movement competence. In the present study, all PLAYfun items demonstrated adequate CITC values, indicating that no item showed a clearly weak relationship with the total score. This finding is also consistent with Longmuir et al. (2018), who emphasized that physical literacy assessment should include observable movement competence that can be evaluated systematically. Compared with self-report instruments, observational measures such as PLAYfun may produce more direct response patterns because scoring is based on observed motor performance rather than on individual perceptions.

Table 4
Item–total correlations and corrected item–total correlations for PLAYself

Item	Pearson item-total *r*	CITC	CITC evaluation
PS1	0.588	0.532	Adequate
PS2	0.554	0.499	Adequate
PS3	0.577	0.520	Adequate
PS4	0.569	0.509	Adequate
PS5	0.282	0.192	Requires further evaluation
PS6	0.307	0.223	Requires further evaluation
PS7	0.314	0.227	Requires further evaluation
PS8	0.365	0.248	Requires further evaluation
PS9	0.681	0.646	Strong
PS10	0.602	0.537	Adequate
PS11	0.539	0.469	Adequate
PS12	0.883	0.865	Very strong
PS13	0.629	0.573	Adequate
PS14	0.811	0.781	Strong
PS15	0.639	0.579	Adequate
PS16	0.821	0.794	Strong
PS17	0.646	0.586	Adequate
PS18	0.912	0.898	Very strong
PS19	0.265	0.230	Requires further evaluation
PS20	0.589	0.549	Adequate
PS21	0.718	0.676	Strong

Source: IBM SPSS Statistics version 26.0

Table 5
Summary of the psychometric characteristics of PLAYfun and PLAYself

Instrument	Number of items	Pearson item–total r	CITC	Cronbach's α	Main finding
PLAYfun	18	0.570–0.893	0.504–0.879	0.955	All items had CITC values > 0.30
PLAYself	21	0.265–0.912	0.192–0.898	0.895	Five items had CITC values < 0.30

The Cronbach's alpha value of 0.955 indicates that the PLAYfun items were highly internally consistent. However, a very high alpha coefficient should not automatically be interpreted as evidence of superior measurement quality. Tavakol and Dennick (2011) noted that alpha values approaching 1.00 may indicate that some items assess highly similar content. Therefore, although the present findings support the internal consistency of PLAYfun, future studies should examine whether each item contributes unique information and whether some degree of item redundancy is present. This issue is particularly important when the instrument is used in educational settings, where assessment efficiency and administration time are practical considerations.

For PLAYself, the Cronbach's alpha value of 0.895 indicates high internal consistency at the scale level. This finding is consistent with Jefferies et al. (2021), who reported favorable psychometric characteristics for PLAYself as a self-reported measure of physical literacy. Nevertheless, five items in the present study—PS5, PS6, PS7, PS8, and PS19—had CITC values below 0.30. This pattern indicates that, although the scale as a whole demonstrated good internal consistency, the contribution of individual items was not uniform. Accordingly, interpretation of PLAYself should not rely solely on the overall alpha coefficient.

This difference may be related to the self-report nature of PLAYself. Students' perceptions of their competence, motivation, and physical activity experiences may be influenced by their understanding of the questions, previous experiences, social environment, and familiarity with the activities described. Whitehead (2010) emphasized that physical literacy includes affective and perceptual components that are inherently individual, which may theoretically lead to greater variability in self-reported responses than in direct assessments of motor performance. In addition, Jefferies et al. (2021) indicated that the interpretation of

Hidayat, Setyawati, & Mukarromah; PELTA 1(1) 2026 8-14 ISSN 3164-7987 (Online)

PLAYself items may vary according to the characteristics of the population being assessed. Therefore, the low CITC values observed for the five items do not necessarily indicate that these items lack value; rather, they may suggest that the item content was less closely aligned with the experiences or understanding of some students in the present context.

Contextual adaptation is another factor that should be considered. Two activities in PLAYself were modified to more closely reflect the experiences of Indonesian students: ice skating was replaced with swimming in rivers or the sea, while cross-country skiing was replaced with cycling on rural or mountainous roads. These modifications were intended to improve the contextual relevance of the items to students' environments and everyday experiences. In theory, contextual adaptation may improve comprehension by making the situations described more familiar to respondents. However, because the adaptation process in the present study did not involve a complete cross-cultural adaptation procedure, such as back-translation, cognitive interviewing, or formal equivalence testing, it cannot yet be established that the replacement activities are fully equivalent to the original items. This limitation may partly explain why several PLAYself items showed weaker relationships with the total score.

Clinical implications

From a practical perspective, the findings suggest that PLAYfun has potential as an assessment tool that can help physical education teachers obtain a structured overview of elementary school students' movement competence. PLAYself may complement this information by providing insight into students' perceptions of their abilities and engagement in physical activity. Using both instruments together may provide a broader perspective on physical literacy because one instrument is based on direct observation, whereas the other relies on self-report. However, for PLAYself in particular, the five items with low CITC values require more detailed examination before the instrument is used to support individual-level decision-making or applied across broader populations.

Study limitations

This study has several limitations. First, the sample consisted of only 71 fifth-grade students from three elementary schools in Pekalongan Regency; therefore, the findings cannot be generalized to all elementary school students in Indonesia. Second, the psychometric analyses were limited to item–total correlations, CITC, and Cronbach's alpha and therefore did not provide evidence regarding factor structure, temporal stability, or inter-rater reliability in the present sample. Third, the contextual adaptation process did not follow a comprehensive cross-cultural adaptation procedure. Fourth, five PLAYself items showed CITC values below 0.30, indicating the need for further evaluation of item content and comprehension in larger and more diverse populations. Future studies should therefore include larger and more heterogeneous samples, conduct Confirmatory Factor Analysis (CFA), assess test–retest reliability for PLAYself, evaluate inter-rater reliability for PLAYfun, and apply more systematic cross-cultural adaptation procedures in accordance with established psychometric instrument-development practices.

Conslusions

This study provides preliminary psychometric evidence regarding item–total relationships and the internal consistency of PLAYfun and PLAYself among elementary school students in the Indonesian context. PLAYfun demonstrated a consistent item–total pattern, with all items showing corrected item–total correlation (CITC) values above 0.30 and very high internal consistency (Cronbach's $\alpha = 0.955$). PLAYself also demonstrated high internal consistency (Cronbach's $\alpha = 0.895$); however, five items—PS5, PS6, PS7, PS8, and PS19—had CITC values below 0.30. These findings indicate that PLAYself showed good overall internal consistency, although several items require further evaluation before the instrument is applied more broadly.

The contextual adaptation of two PLAYself activities—replacing ice skating with swimming in rivers or the sea and cross-country skiing with cycling on rural or mountainous roads—represented an initial effort to improve the relevance of the instrument to the physical activity experiences of Indonesian students. However, the present study did not include comprehensive cross-cultural validation or construct validation; therefore, the psychometric equivalence of the adapted version with the original instrument remains to be established.

From a practical perspective, PLAYfun may be considered as an observational instrument for obtaining information on students' movement competence, while PLAYself may provide complementary information regarding students' perceptions of their physical literacy. Nevertheless, the present findings should be regarded as a basis for further instrument development and testing rather than as evidence of national validation of the PLAY instruments. Future studies should involve larger and more diverse samples, examine factor structure using Confirmatory Factor Analysis (CFA), evaluate test–retest reliability for PLAYself and inter-rater reliability for PLAYfun, and adopt more systematic cross-cultural adaptation procedures. These steps would strengthen the evidence supporting the use of PLAYfun and PLAYself for assessing physical literacy among Indonesian elementary school students.

Acknowledgments

The authors gratefully acknowledge the Pusat Pembiayaan Asesmen Pendidikan Tinggi (PPAPT) and the Lembaga Pengelola Dana Pendidikan (LPDP) for providing financial support for this research.

Author contributions

Conceptualization: R.H., H.S., and S.B.M.; Methodology: R.H., H.S., and S.B.M.; Investigation: R.H.; Data curation: R.H.; Formal analysis: R.H. and H.S.; Project administration: R.H.; Supervision: H.S. and S.B.M.; Visualization: R.H.; Writing – original draft: R.H.; Writing – review & editing: R.H., H.S., and S.B.M. All authors have read and approved the final version of the manuscript and agree to be accountable for the integrity and accuracy of the work.

Ethical statement

This study was reviewed by the Health Research Ethics Committee, Faculty of Medicine, Universitas Negeri Semarang, Indonesia, and was granted an Ethical Exemption under reference No. 813/KEPK/FK/KLE/2025. The study was conducted in accordance with applicable ethical principles for research involving human participants. Permission to conduct the research was also obtained from the participating schools prior to data collection. Participation was voluntary, and the confidentiality and anonymity of all participants were maintained throughout the study.

Informed consent

Written informed consent was obtained from the parents or legal guardians of all participating students prior to data collection. In addition, all students received an age-appropriate explanation of the study procedures and provided their voluntary assent to participate. Both the students and their parents or legal guardians were informed that participation was voluntary and that the students could withdraw from the study at any time without penalty. All collected data were treated confidentially, and no personally identifiable information was included in the analysis or reported in the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely to improve language, grammar, and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the accuracy, integrity, and content of the publication.

Funding

This research was supported by the Lembaga Pengelola Dana Pendidikan (LPDP) through the Pusat Pembiayaan Asesmen Pendidikan Tinggi (PPAPT). The funders had no role in the study design, data collection, data analysis, interpretation of the findings, or preparation of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to appropriate ethical and privacy considerations.

References

Arikunto, S. (2010). *Research Design. Pendekatan Metode Kualitatif*, Al Fabet, Bandung, 22.

Barnett, L. M., Jerebine, A., Keegan, R., Watson-Mackie, K., Arundell, L., Ridgers, N. D., Salmon, J., & Dudley, D. (2023). Validity, Reliability, and Feasibility of Physical Literacy Assessments Designed for School Children: A Systematic Review. *Sports Medicine, 53*(10), 1905–1929. https://doi.org/10.1007/s40279-023-01867-4

Cairney, J., Dudley, D., Kwan, M., Bulten, R., & Kriellaars, D. (2019). 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. T., Di Cristofaro, N. A., Cairney, J., Bray, S. R., MacDonald, M. J., & Timmons, B. W. (2020). Physical Literacy, Physical Activity, and Health Indicators in School-Age Children. *International Journal of Environmental Research and Public Health, 17*(15), 5367. https://doi.org/10.3390/ijerph17155367

Etz, R. S., Zyzanski, S. J., Gonzalez, M. M., Reves, S. R., O'Neal, J. P., & Stange, K. C. (2019). A New Comprehensive Measure of High-Value Aspects of Primary Care. *The Annals of Family Medicine, 17*(3), 221–230. https://doi.org/10.1370/afm.2393

Fu, T., Zhang, D., Wang, W., Geng, H., Lv, Y., Shen, R., & Bu, T. (2022). Functional Training Focused on Motor Development Enhances Gross Motor, Physical Fitness, and Sensory Integration in 5–6-Year-Old Healthy Chinese Children. *Frontiers in Pediatrics, 10*. https://doi.org/10.3389/fped.2022.936799

Giblin, S., Collins, D., & Button, C. (2014). Physical Literacy: Importance, Assessment and Future Directions. *Sports Medicine, 44*(9), 1177–1184. https://doi.org/10.1007/s40279-014-0205-7

Gunnell, K. E., Longmuir, P. E., Barnes, J. D., Belanger, K., & Tremblay, M. S. (2018). Refining the Canadian Assessment of Physical Literacy based on theory and factor analyses. *BMC Public Health, 18*(S2), 1044. https://doi.org/10.1186/s12889-018-5899-2

Jefferies, P., Bremer, E., Kozera, T., Cairney, J., & Kriellaars, D. (2021). Psychometric properties and construct validity of PLAYself: a self-reported measure of physical literacy for children and youth. *Applied Physiology, Nutrition, and Metabolism, 46*(6), 579–588. https://doi.org/10.1139/apnm-2020-0410

Knisel, E., Bremer, M., Nałęcz, H., Wascher, L., & Laudańska-Krzemińska, I. (2024). Validation of The Canadian Assessment of Physical Literacy – CAPL-2 Questionnaire for German and Polish School Children. Physical Culture and Sport. *Studies and Research, 104*(1), 1–14. https://doi.org/10.2478/pcssr-2024-0014

Kriellaars, D., & Robillard, B. (2013). *PLAYself: Physical Literacy Assessment for Youth.* Canadian Sport for Life, Victoria, BC, Canada.

Longmuir, P. E., Gunnell, K. E., Barnes, J. D., Belanger, K., Leduc, G., Woodruff, S. J., & Tremblay, M. S. (2018). Canadian Assessment of Physical Literacy Second Edition: a streamlined assessment of the capacity for physical activity among children 8 to 12 years of age. *BMC Public Health, 18*(S2), 1047. https://doi.org/10.1186/s12889-018-5902-y

Mentzel, T. Q., Lieverse, R., Levens, A., Mentzel, C. L., Tenback, D. E., Bakker, P. R., Daanen, H. A. M., & van Harten, P. N. (2016). Reliability and validity of an instrument for the assessment of bradykinesia. *Psychiatry Research, 238*, 189–195. https://doi.org/10.1016/j.psychres.2016.02.011

Saptono, A. (2017). Development Instruments Through Confirmatory Factor Analysis (CFA) in Appropriate Intensity Assessment. *Dinamika Pendidikan, 12*(1), 13–19. https://doi.org/10.15294/dp.v12i1.10578

Scott, J. J., Hill, S., Barwood, D., & Penney, D. (2021). Physical literacy and policy alignment in sport and education in Australia. *European Physical Education Review, 27*(2), 328–347. https://doi.org/10.1177/1356336X20947434

Shearer, C., Goss, H. R., Boddy, L. M., Knowles, Z. R., Durden-Myers, E. J., & Foweather, L. (2021a). Assessments Related to the Physical, Affective and Cognitive Domains of Physical Literacy Amongst Children Aged 7–11.9 Years: A Systematic Review. *Sports Medicine - Open, 7*(1), 37. https://doi.org/10.1186/s40798-021-00324-8

Shearer, C., Goss, H. R., Boddy, L. M., Knowles, Z. R., Durden-Myers, E. J., & Foweather, L. (2021b). Assessments Related to the Physical, Affective and Cognitive Domains of Physical Literacy Amongst Children Aged 7–11.9 Years: A Systematic Review. *Sports Medicine - Open, 7*(1). https://doi.org/10.1186/s40798-021-00324-8

Stearns, J. A., Wohlers, B., McHugh, T.-L. F., Kuzik, N., & Spence, J. C. (2019). Reliability and Validity of the PLAY fun Tool with Children and Youth in Northern Canada. *Measurement in Physical Education and Exercise Science, 23*(1), 47–57. https://doi.org/10.1080/1091367X.2018.1500368

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, R & D*. CV Alfabeta.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Whitehead, M. E. (2010). *Physical Literacy Throughout the Lifecourse*. In U. Richard Bailey, University of Birmingham & Routledge's (Eds.), Taylor & Francis e-Library (first, p. 45). Routledge.