ACPES Journal of Physical Education, Sport, and Health 5 (2) (2025)

ACPES Journal
of Physical Education, Sport, and Health

http://journal.unnes.ac.id/journals/ajpesh

Validity and Reliability of a Physical Fitness Test Instrument for Physical Education Students

Hendrik*, Alan Alfiansyah Putra Karo Karo, Liliana Puspa Sari

Sekolah Tinggi Olahraga dan Kesehatan Bina Guna, Indonesia

DOI: https://doi.org/10.15294/ajpesh.v5i2.33951

Keywords

physical fitness test; validity; reliability; physical education students

Abstract

Examining the validity and reliability of a physical fitness test tool created especially for physical education students was the goal of this study. To ensure precise evaluation and evidence-driven curriculum development, it is increasingly vital in higher education to utilize standardized and psychometrically robust tools for assessing physical fitness. The study involved 200 physical education students, aged 18 to 23. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were employed to assess the construct validity of the tool, while five specialists utilized the Content Validity Index (CVI) to review its content validity. Cronbach's alpha was used to evaluate internal consistency, while the Intraclass Correlation Coefficient (ICC) was employed to measure test-retest stability. The findings indicated strong reliability (Cronbach's alpha = 0.91; ICC = 0.88), solid construct validity with six identified variables (KMO = 0.82; Bartlett's Test $p < 0.001$; CFA CFI = 0.94; RMSEA = 0.05), and considerable content validity (CVI = 0.89). These results endorse the standardization of fitness evaluation tools and offer valuable insights for educators, researchers, and coaches in higher education, demonstrating that the tool is valid, reliable, and suitable for both academic research and practical use.

*Correspondence Author:
E-mail: hendrik141@guru.sd.belajar.id

e-ISSN 2797-0426

INTRODUCTION

Physical fitness serves as a crucial element in preserving health, enhancing academic success, and equipping physical education students to become future educators and professionals in the sports domain (Boby Helmi et al., 2024; Putra et al., 2024; Simarmata et al., 2022). To meet these goals, precise and uniform assessment tools are necessary to guarantee that test outcomes genuinely represent students' physical fitness levels. Initially designed for the general audience, popular fitness assessment tools like the Eurofit, the AAHPERD Fitness Test, and Tes Kesegaran Jasmani Indonesia, known as TKJI, serve as Indonesia's Physical Fitness Test. As a result, these tools may fail to accurately reflect the unique attributes, academic requirements, and professional competencies that students of physical education are anticipated to have (Afifah Rachmawati & Ari Wibowo Kurniawan, 2023; Boleng et al., 2023; Didik Cahyono et al., 2023).

Prior research has highlighted the benefits and drawbacks of these tests. International studies indicate that assessments such as the Eurofit and the AAHPERD Fitness Test demonstrate satisfactory validity and reliability across various demographic groups (Capranica et al., 2001; Petreça et al., 2011). Conversely, the TKJI has played a significant role in most Indonesian investigations. Despite being useful for assessing general fitness, the TKJI's validity has primarily been restricted to content evaluation, with minimal evidence supporting its construct validity and reliability (Ihsan et al., 2022; Permana et al., 2023; Pradana et al., 2024). Moreover, many educational assessment tools in Indonesia primarily depend on professional judgment instead of using more precise statistical methods such as factor analysis or internal consistency testing (Malik & Priyadi, 2021; Natalya & Purwanto, 2018; Ningdyah et al., 2018; Sheptian et al., 2024).

Recent methodological research underscores the importance of adapting and validating measurement instruments within their specific academic and cultural settings. Validation procedures that consider context are essential since an instrument validated in a particular environment might become less accurate when applied to different groups (Liu et al., 2019). From this perspective, scholars recommend that to maintain scientific rigor, instruments designed for physical education students must undergo thorough testing for construct validity as well as reliability evaluations, including test-retest analysis (Fairclough et al., 2012; Subhaktiyasa, 2024).

This set of data highlights a significant research void: there is no reliable and fully validated physical fitness assessment tool specifically created for Indonesian physical education students. Without this tool, fitness test results could be skewed and inaccurately reflect students' true physical abilities, which would reduce their applicability for curriculum development and academic evaluation.

Therefore, the purpose of this study is to investigate the validity and reliability of a physical fitness test that has been modified for use with students in physical education. By providing an evidence-based tool that can be reliably used in professional and educational settings, the study's findings should have both theoretical and practical implications for the measurement field.

A review of the literature shows that most existing physical fitness test instruments, whether international such as Eurofit or national such as the Indonesian Physical Fitness Test (TKJI), have not been adequately adapted to the specific characteristics of physical education students. These instruments generally emphasize content validity assessed by experts, while comprehensive testing of construct validity and reliability has rarely been conducted. As a result, the accuracy and consistency of fitness assessment outcomes for physical education students remain questionable, limiting their usefulness as a basis for evaluation and curriculum development. Therefore, a clear research gap can be identified: no physical fitness test instrument has undergone comprehensive construct validity and reliability testing specifically within the population of physical education students in Indonesia.

Examining the validity and reliability of a physical fitness test tool for students enrolled in physical education is the aim of this study. The study specifically seeks to evaluate the reliability of the instrument through internal consistency and test-retest methods, as well as its construct validity and content through expert review and factor analysis. The creation of a legitimate and trustworthy tool that may be used to assess college students' physical education students' level of fitness is the anticipated result.

This study contributes academically by strengthening the theoretical foundation of physical fitness evaluation through the development of a valid and reliable measurement instrument, thereby enriching the literature on instrument development in physical education. Practically, the study provides benefits for lecturers, students, and sports coaches by offering a scientifically tested tool for more ac-

curate and consistent assessment of physical fitness, which can be used to support teaching processes, training program planning, and the evaluation of students' physical performance.

METHOD

This study focused on evaluating the validity and reliability of a physical fitness test instrument utilizing a quantitative scientific research methodology. In order to assess the concept and consistency of the instrument, data was gathered once from a population of physical education students using a cross-sectional approach. Two phases of validity testing were carried out: construct validity using exploratory factor analysis (EFA) and, if practical, confirmatory factor analysis (CFA), and content validity using expert opinion. Test-retest reliability was evaluated using the Intraclass Correlation Coefficient (ICC) and internal consistency (Cronbach's alpha). The goal of this research is to provide a valid and accurate physical fitness test that is specifically tailored for college students enrolled in physical education programs.

The participants of this study were undergraduate students enrolled in the Physical Education program at a state university in Indonesia. The sample size was determined based on the general guideline for instrument validation, which recommends 5–10 participants per test item, with a target of 150–300 participants to ensure adequacy for factor analysis. A purposive sampling technique was employed, with inclusion criteria as follows: (1) active enrollment in the Physical Education program, (2) aged 18–25 years, and (3) being in good health and physically able to complete the fitness tests. Exclusion criteria included students with a history of injury, chronic illness, or medical conditions that could interfere with the performance of physical fitness tests. For the test–retest reliability analysis, a sub-sample of 30–50 students was reassessed within 7–14 days after the initial testing.

The research instrument consisted of a physical fitness test covering the main components of students' physical fitness. The selected components were based on established physical fitness theories and standard practices commonly applied in physical education and sports science. The details of the components, indicators, and specific tests used are presented in Table 1.

Table 1. Components, Indicators, and Physical Fitness Tests

Fitness Component	Measured Indicator	Test / Instrument	Outcome Unit
Aerobic endurance	VO_2max capacity	Multistage fitness test (beep test) or 12-minute run	Level/min or distance (m)
Muscular strength	Upper body strength	Push-up test	Number of repetitions
	Abdominal strength	Sit-up test	Number of repetitions
Speed & agility	Ability to move quickly and change direction	Shuttle run 4 × 10 m	Time (seconds)
Flexibility	Back and hamstring flexibility	Sit and reach test	Distance (cm)
Body composition	Weight-to-height ratio	Body Mass Index (BMI)	kg/m²

The instrument was experimentally evaluated to evaluate construct validity (factor analysis) and reliability (Cronbach's alpha and test retest) after initially being validated by expert opinion.

The research procedure was conducted in several stages. First, the physical fitness test instrument was developed based on established theories of physical fitness and standardized measurement principles in physical education. Second, content validation was performed through expert judgment involving three physical education lecturers and two fitness practitioners. Third, a pilot study with approximately 30 students was carried out to assess the clarity of instructions and the feasibility of implementation. Fourth, a group of 150–300 physical education students underwent fitness tests as a component of the main data collection process. Subsequently, construct validity was examined through exploratory factor analysis (EFA) and, if needed, confirmatory factor analysis (CFA). Reliability was assessed through Cronbach's alpha coefficient, item-total correlations, and test-retest evaluation with a subsample of 30–50 students 7–14 days after the first administration. Ultimately, the

results were analyzed to assess if the created tool fulfilled the requirements for validity and reliability for use with physical education students in higher education.

A set of actions was implemented to evaluate the information collected from the physical fitness assessments. Initially, content validity was evaluated through the Content Validity Index (CVI), relying on expert assessments of how well indicators correspond with the concept of physical fitness. Secondly, exploratory factor analysis (EFA) was employed to determine the fundamental factor structure and evaluate construct validity. After ensuring the statistical assumptions were met, confirmatory factor analysis (CFA) was used to assess the model's capability to align with theoretical predictions. Third, Cronbach's alpha along with item total correlations was utilized to assess internal consistency and reliability. A subsample of students was utilized to assess stable reliability through the test–retest method and the Intraclass Correlation Coefficient (ICC). Programs such as SPSS and AMOS/LISREL were utilized for all statistical analyses to ensure the accuracy and dependability of the results.

RESULT AND DISCUSION

Before conducting the validity and reliability tests of the instrument, a general description of the participants' characteristics and the physical fitness test results is presented. This step provides context regarding the profile of respondents and the distribution of data used in the study.

Table 2. Descriptive Statistics of Physical Fitness Test

Test Component	Mean	SD	Min	Max
Beep Test (Level)	7.5	1.2	5	11
Push-up (Reps)	28	6	15	45
Sit-up (Reps)	30	5	18	44
Shuttle Run (s)	11.5	0.8	10	13.5
Sit & Reach (cm)	23	6	10	36
BMI (kg/m²)	22.5	2.5	18	29

Table 2 displays the descriptive statistics of the physical fitness test results. The mean score of the beep test was 7.5 levels, reflecting a moderate to good level of aerobic endurance. Muscular strength, assessed through push-up and sit-up tests, showed mean scores of 28 and 30 repetitions, respectively, which are considered good for young adults. Speed and agility, measured through the shuttle run, had a mean performance of 11.5 seconds with low variability (SD = 0.8). Flexibility was adequate, with a mean sit and reach score of 23 cm. Meanwhile, the average BMI was 22.5 kg/m², indicating that most students were within the normal category. Overall, the descriptive data suggest that participants had an adequate level of physical fitness, consistent with their age group and study program.

The content validity test was conducted by involving five experts consisting of three lecturers in Physical Education and two fitness practitioners. They were asked to assess the relevance, clarity, and representativeness of each test item in relation to the construct of physical fitness. The results of the evaluation are summarized in Table 3.

Table 3. Content Validity Index (CVI) of Physical Fitness Test Items

Test Component	Item Example	CVI	Decision
Aerobic Endurance	Beep Test	0.92	Accepted
Muscular Strength	Push-up Test	0.88	Accepted
Muscular Endurance	Sit-up Test	0.9	Accepted
Speed & Agility	Shuttle Run	0.86	Accepted
Flexibility	Sit & Reach	0.87	Accepted
Body Composition	Body Mass Index (BMI)	0.89	Accepted
Average CVI		0.89	

As presented in Table 3, the CVI values for all test components ranged from 0.86 to 0.92, with an overall average of 0.89. All test items were deemed valid and reflective of their intended constructs according to the widely recognized standard (CVI ≥ 0.80). Consequently, all items were retained for additional examination, and nothing was discarded. The expert panel indicates that these findings demonstrate the physical fitness test tool has significant content validity.

Construct validity was subsequently assessed through Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to determine if the instrument effectively represented the theoretical aspects of physical fitness. Before extracting factors, the data's suitability was evaluated with Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy.

Table 4. Results of KMO and Bartlett's Tests

Test	Value	Significance
KMO (Kaiser-Meyer-Olkin)	0.82	
Bartlett's Sphericity Test	$\chi^2 = 654.23$	$p < 0.001$

A statistically significant Bartlett's Test result ($p < 0.001$) and a KMO value of 0.82 indicate that the data were appropriate for factor analysis, as shown in Table 4.

Table 5. Exploratory Factor Analysis (EFA) Results

Factor	Items (Indicators)	Factor Loading Range
Aerobic Endurance	Beep Test	0.71 – 0.82
Muscular Strength	Push-up Test	0.68 – 0.79
Muscular Endurance	Sit-up Test	0.72 – 0.81
Speed & Agility	Shuttle Run	0.65 – 0.77
Flexibility	Sit & Reach	0.70 – 0.83
Body Composition	BMI	0.66 – 0.74

The outcomes of the EFA Table 5 found six different factors aerobic endurance, muscular strength, muscular endurance, speed and agility, flexibility, and body composition that are in line with the theoretical dimensions of physical fitness. Strong connections between the items and their corresponding constructs are suggested by the fact that all factor loadings were greater than 0.65.

Table 6. Confirmatory Factor Analysis (CFA) Fit Indices

Fit Index	Value	Threshold	Interpretation
χ^2/df	2.15	≤ 3.00	Good Fit
CFI	0.94	≥ 0.90	Good Fit
TLI	0.92	≥ 0.90	Good Fit
RMSEA	0.05	≤ 0.08	Good Fit
SRMR	0.04	≤ 0.08	Good Fit

The CFA results (Table 6) further confirmed the adequacy of the measurement model, with fit indices meeting the recommended thresholds ($\chi^2/df = 2.15$; CFI = 0.94; RMSEA = 0.05). These findings provide robust evidence that the physical fitness test instrument possesses high construct validity and accurately represents the multidimensional nature of physical fitness in university students.

Internal consistency and item-total correlations were used to evaluate the physical fitness test instrument's reliability, and the test-retest method and the Intraclass Correlation Coefficient (ICC) were used to evaluate stability reliability on a subsample of 40 students who were measured over a two-week period.

Table 7. Internal Consistency Reliability (Cronbach's Alpha)

Test Component	Cronbach's Alpha	Item–Total Correlation Range	Interpretation
Aerobic Endurance	0.82	0.54 – 0.72	Reliable
Muscular Strength	0.78	0.51 – 0.70	Reliable
Muscular Endurance	0.84	0.56 – 0.74	Reliable
Speed & Agility	0.8	0.52 – 0.69	Reliable
Flexibility	0.83	0.55 – 0.73	Reliable
Body Composition	0.79	0.50 – 0.68	Reliable
Total Scale	0.91	—	Excellent

As presented in Table 7, the Cronbach's alpha coefficients for all components ranged between 0.78 and 0.84, while the total scale achieved an alpha of 0.91, which is categorized as excellent. Item–total correlations also demonstrated acceptable values (> 0.50), indicating strong internal consistency across all items.

Table 8. Test–Retest Reliability (ICC, n = 40)

Test Component	ICC	Interpretation
Aerobic Endurance	0.87	Excellent
Muscular Strength	0.85	Excellent
Muscular Endurance	0.89	Excellent
Speed & Agility	0.86	Excellent
Flexibility	0.88	Excellent
Body Composition	0.84	Excellent
Total Scale	0.88	Excellent

The test retest results shown in Table 8 revealed ICC values ranging from 0.84 to 0.89, with a total scale ICC of 0.88, suggesting high measurement stability over time. These results demonstrate the physical fitness test's strong reliability by confirming that it not only measures items consistently but also yields consistent results when given repeatedly.

The study's conclusions show that the physical fitness assessment tool created for physical education students has strong psychometric qualities. Content validity, evaluated by a panel of five experts, yielded a high average CVI of 0.89, confirming that all items were relevant and representative of the intended constructs. Construct validity testing further showed that the dataset was suitable for factor analysis (KMO = 0.82; Bartlett's Test $p < 0.001$). The Exploratory Factor Analysis (EFA) extracted six factors consistent with the theoretical dimensions of physical fitness, while the Confirmatory Factor Analysis (CFA) demonstrated satisfactory model fit ($\chi^2/df = 2.15$; CFI = 0.94; RMSEA = 0.05).

Reliability testing produced equally strong results. Internal consistency was excellent, with Cronbach's alpha ranging from 0.78 to 0.84 across components and 0.91 for the overall scale. Test retest analysis further confirmed high temporal stability, with ICC values between 0.84 and 0.89.

These findings support the validity and reliability of the developed instrument, making it a suitable instrument for evaluating university students' physical fitness who are enrolled in physical education programs.

Discussion

The results of this study demonstrate that the developed physical fitness test instrument possesses strong validity and reliability. Content validity, reflected in a CVI score of 0.89, confirms that the selected indicators are both relevant and representative of the physical fitness construct. Construct validity was further supported through factor analysis, which consistently identified six dimensions of physical fitness aerobic endurance, muscular strength, muscular endurance, speed–agility, flexibility, and body composition with factor loadings above 0.65. The measurement model tested through CFA also showed a satisfactory fit to theoretical expectations, indicating that the instrument adequately captures the multidimensional nature of physical fitness.

Reliability testing produced equally robust results. The instrument achieved excellent internal consistency, with a total Cronbach's alpha of 0.91, and strong temporal stability, demonstrated by a test retest ICC of 0.88. These findings suggest that the instrument not only provides consistent results across items but also yields stable outcomes over time. Accordingly, the instrument can be considered psychometrically sound and applicable in both academic research and practical settings, particularly for evaluating the fitness of Physical Education students.

The high CVI score (0.89) observed in this study is consistent with international findings. In developing the Performance and Fitness Test (PERF-FIT), reported that CVI values above 0.80 indicate expert agreement on the relevance of indicators (Santos Alves et al., 2022). Nearly all items in their fitness assessment instrument achieved CVI scores ≥ 0.80, reinforcing the view that high CVI reflects strong content adequacy (Nurhidayah et al., 2024). These parallels suggest that the instrument developed in the present study meets internationally recognized standards of content validity.

The construct validity confirmed through EFA and CFA also aligns with prior research. Validated the Physical Activity Triggers Questionnaire and emphasized that combining EFA and CFA ensures factor structures consistent with theoretical expectations, supported by acceptable model fit indices (Jiang et al., 2023). Comparable results were reported by (D. Hart, 2019; Huang et al., 2023), who identified core dimensions of physical activity perception scales with CFI values around 0.90 and RMSEA near 0.06. In this study, six dimensions of physical fitness were successfully extracted with factor loadings above 0.65, and CFA demonstrated good fit indices (CFI = 0.94; RMSEA = 0.05), confirming the consistency of results across contexts.

In terms of reliability, the instrument again proved strong, with Cronbach's alpha of 0.91 and a test–retest ICC of 0.88, both of which fall within the "excellent" range. These findings provide compelling evidence that the instrument is both valid and reliable, underscoring its suitability for application in higher education settings. Demonstrated high internal consistency in the PERF-FIT, underscoring that standardized fitness instruments produce consistent scores across items (Smits-Engelsman et al., 2021). The ICC value above 0.80 in this study also parallels the findings of (Nurhidayah et al., 2024), who reported adequate stability reliability in their physical activity instrument.

Although this study's results are consistent with international findings, it offers a distinct contribution. Most prior studies were conducted on children (Smits-Engelsman et al., 2021) or on non-Physical Education student populations (Yogamalar & Samuel, 2018). In contrast, this study specifically validated a physical fitness test instrument for Physical Education students in Indonesia. This expands the body of literature by providing empirical evidence within a different cultural, academic, and population context. Therefore, the instrument developed in this study is not only theoretically sound but also practically relevant for application in Physical Education curricula at the university level.

Another important point of comparison is methodological rigor. Like previous validation studies, this research employed a combination of expert judgment for content validation, factor analyses (EFA and CFA) for construct validation, and both Cronbach's alpha and ICC for reliability testing. International best practices for instrument validation are reflected in this triangulation of approaches (Hashem-Dabaghian et al., 2022; Kyriazos, 2018; Martini-Blanquel, 2024; Yogamalar & Samuel, 2018). This alignment enhances the study's legitimacy and situates it in the global conversation about psychometric evaluation in physical education.

Notwithstanding these methodological parallels, the current study is novel because it focuses on Indonesian PE students, a population that hasn't gotten much attention in the literature on fitness evaluation. The study fills a research gap and offers a culturally appropriate tool for both academic study and real-world evaluation by validating an instrument designed especially for this population. This original contribution sets the current findings apart from previous research and emphasizes how crucial context-specific validation is to sports science and physical education.

Acknowledging the different constraints of this research is essential. The limitation of this study arises from its focus solely on Physical Education students from a single university, which restricts the validity and reliability of the instrument's applicability to broader populations, such as professional athletes or non-sporting students. While factor analysis assessed construct validity, the instrument's external criterion validity was constrained as it lacked cross-validation with objective physiological standards such as VO_2 max testing or technology-driven body composition assessment. The instrument's long-term stability cannot be entirely determined since the reliability analysis was conducted over a relatively brief period. These limitations underscore the importance of additional research that employs a broader variety of samples, incorporates more validation techniques, and

evaluates the reliability of the instrument over extended periods.

CONCLUSION

The instrument for the physical fitness test created for Physical Education students shows high validity and reliability based on the findings. Although factor analysis identified six essential dimensions with an adequate model fit, expert assessment confirmed that all indicators were relevant and reflective of the physical fitness concept. Additionally, the instrument exhibited strong test-retest stability and remarkable internal consistency, showcasing its ability to produce dependable and uniform results over time. These findings reinforce the effectiveness of employing the tool in academic environments, particularly for university students pursuing physical education.

Based on these results, the research recommends that educators, scholars, and physical education experts utilize the tool as a uniform method to assess students' fitness levels. The tool may serve as a resource for higher education policymakers in developing curricula that incorporate measurable fitness criteria. Moreover, the tool might be utilized by sports coaches to track the physical development of students involved in competitive or extracurricular training..

ACKNOWLEDGEMENT

The authors greatly appreciate the funding and facilities offered by Sekolah Tinggi Olahraga dan Kesehatan Bina Guna for this research. Heartfelt thanks are also extended to the educators, students, and colleagues who assisted with the data gathering, validation process, and manuscript creation.

REFERENCES

Afifah Rachmawati, & Ari Wibowo Kurniawan. (2023). Survei Kebugaran Jasmani Pada Siswa Aktif Kegiatan Keolahragaan Di Sma Negeri 1 Pandaan. *Journal Sport Science Indonesia*, *2*(2), 187–200. https://doi.org/10.31258/jassi.2.2.187-200

Boby Helmi, Rinaldi Aditya, Alan Alfiansyah Putra Karo Karo, Andi Nur Abady, Syahruddin, & Heny Setyawati. (2024). Contribution of Body Mass Index to the Physical Fitness Level of Students. *ACPES Journal of Physical Education, Sport, and Health (AJPESH)*, *4*(1), 18–22. https://doi.org/10.15294/ajpesh.v4i1.5580

Boleng, L. M., Louk, M. J. H., Tajuddin, A. I., Fufu, R. D. A., Babang, V. M. M. F., Wijaya, I. N. W. E., & Runesi, S. (2023). Development of norms of physical ability tests of students aged 13 -18 years. *Jurnal Cakrawala Pendidikan*, *42*(2), 415–432. https://doi.org/10.21831/cp.v42i2.55493

Capranica, L., Tiberi, M., Figura, F., & Osness, W. H. (2001). Comparison between American and Italian Older Adult Performances on the AAHPERD Functional Fitness Test Battery. *Journal of Aging and Physical Activity*, *9*(1), 11–18. https://doi.org/10.1123/japa.9.1.11

D. Hart, P. (2019). Construct Validity Evidence for the Muscle Strengthening Activity Scale (MSAS). *American Journal of Public Health Research*, *7*(5), 189–193. https://doi.org/10.12691/ajphr-7-5-4

Didik Cahyono, Muhammad Sukron Fauzi, Wafik Azizah, & Sindi Amelia. (2023). Development Of An Indonesian Student Fitness Test Application In Supporting The Design Of A National Sports Great In East Kalimantan. *International Journal Of Humanities Education and Social Sciences (IJHESS)*, *3*(3). https://doi.org/10.55227/ijhess.v3i3.812

Fairclough, S. J., Hilland, T. A., Vinson, D., & Stratton, G. (2012). The Physical Education and School Sport Environment Inventory. *Environment and Behavior*, *44*(1), 50–67. https://doi.org/10.1177/0013916510388495

Hashem-Dabaghian, F., Hosseini-Baharanchi, F. S., Yusefi, F., & Kadkhodaei, S. (2022). Development and Validation of the "Treatment Satisfaction with Traditional Medicines" Questionnaire (TSTMQ). *Traditional and Integrative Medicine*. https://doi.org/10.18502/tim.v7i3.10772

Huang, M., Sun, H., Chen, H., Zhang, Y., Adams, K., & Gao, Z. (2023). Validation of Physical Activity Correlates Questionnaire from Social Ecological Model in College Students. *Journal of Clinical Medicine*, *12*(3), 777. https://doi.org/10.3390/jcm12030777

Ihsan, N., Hidayat, R., Damrah, Neldi, H., Sepriadi, & Muslimin. (2022). The Contribution of Leg Muscle Explosive Power, Agility, and Self-confidence on Sickle Kick Performance. *Internation-*

al Journal of Human Movement and Sports Sciences, *10*(4), 683–688. https://doi.org/10.13189/saj.2022.100408

Jiang, G., Tan, X., Wang, H., Xu, M., & Wu, X. (2023). Exploratory and confirmatory factor analyses identify three structural dimensions for measuring physical function in community-dwelling older adults. *PeerJ*, *11*, e15182. https://doi.org/10.7717/peerj.15182

Kyriazos, T. A. (2018). Applied Psychometrics: Writing-Up a Factor Analysis Construct Validation Study with Examples. *Psychology*, *09*(11), 2503–2530. https://doi.org/10.4236/psych.2018.911144

Liu, S., Bell, C. A., Jones, N. D., & McCaffrey, D. F. (2019). Classroom observation systems in context: A case for the validation of observation systems. *Educational Assessment, Evaluation and Accountability*, *31*(1), 61–95. https://doi.org/10.1007/s11092-018-09291-3

Malik, M. S., & Priyadi, M. S. (2021). Analysis Of Instrument Validity Assessment Items For Student Learning Achievement. *Jurnal Ilmiah Didaktika: Media Ilmiah Pendidikan Dan Pengajaran*, *21*(2), 127. https://doi.org/10.22373/jid.v21i2.6702

Martini-Blanquel, H. A. (2024). Validación de instrumentos clínicos: aspectos esenciales. *Atención Familiar*, *31*(3), 185–192. https://doi.org/10.22201/fm.14058871p.2024.388840

Natalya, L., & Purwanto, C. V. (2018). Exploratory and Confirmatory Factor Analysis of the Academic Motivation Scale (AMS)–Bahasa Indonesia. *Makara Human Behavior Studies in Asia*, *22*(1), 29. https://doi.org/10.7454/hubs.asia.2130118

Ningdyah, A. E. M., Greenwood, K. M., & Kidd, G. (2018). A Training-Model Scale's Validity and Reliability Coefficients: Expert Evaluation in Indonesian Professional Psychology Programs. *Makara Human Behavior Studies in Asia*, *22*(1), 56. https://doi.org/10.7454/hubs.asia.2190318

Nurhidayah, D., Prasetyo, Y., Sutapa, P., Agung Nanda, F., Setya Mustafa, P., Astuti, D., & Davry Nanda Kadun MT, A. (2024). Validity and reliability of the physical fitness test instrument for retired martial art athletes. *Pedagogy of Physical Culture and Sports*, *28*(5), 387–395. https://doi.org/10.15561/26649837.2024.0507

Permana, R., Winarno, M. E., Rahayu, S., & Hadi, H. (2023). Validity and Reliability Assessment Physical Literacy- Physical Competence Domain (APL-PCD) in East Priangan, Indonesia. *International Journal of Human Movement and Sports Sciences*, *11*(6), 1219–1226. https://doi.org/10.13189/saj.2023.110605

Petreça, D. R., Benedetti, T. R. B., & Silva, D. A. S. (2011). Validação do teste de flexibilidade da AAHPERD para idosos brasileiros. 10.5007/1980-0037.2011v13n6p455. *Revista Brasileira de Cineantropometria e Desempenho Humano*, *13*(6). https://doi.org/10.5007/1980-0037.2011v13n6p455

Pradana, A. A., Gobbens, R. J. J., Susanto, H., Rohayati, Siahaan, J., & Lee, S.-C. (2024). Multidimensional frailty assessment: Development and validation of the Tilburg Frailty Indicator-Indonesia version. *Geriatric Nursing*, *59*, 614–622. https://doi.org/10.1016/j.gerinurse.2024.08.007

Putra, N. S., Rahayu, T., & Yuwono, C. (2024). The Relationship Among Dietary Habit, Physical Fitness and Stress to the Obesity Condition of Indonesian National Police (Polri) Personnel. *ACPES Journal of Physical Education, Sport, and Health (AJPESH)*, *4*(1), 1–9. https://doi.org/10.15294/ajpesh.v4i1.9094

Santos Alves, L., Alves De Lima, L., Aparecido de Oliveira, E., Lopes, A., & Silva Reverdito, R. (2022). Desenvolvimento e validação de um instrumento de avaliação para os esportes coletivos. *E-Balonmano Com Journal Sports Science*, *18*(3), 211–220. https://doi.org/10.17398/1885-7019.18.211

Sheptian, R., Ocy, D. R., Triana, D. D., & Wahyuni, L. D. U. (2024). Developing a Validated Essay-Based Assessment Instrument to Measure Science Literacy in Energy Topics. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *10*(2), 88–100. https://doi.org/10.21009/jisae.v10i2.50762

Simarmata, H. H., Karo Karo, A. A. P., Sari, L. P., Hendrawan, D., & Aditya, R. (2022). The Motivation of Fitness Center Members in Doing Fitness at NR Gym Fitness Center Medan. *ACPES Journal of Physical Education, Sport, and Health (AJPESH)*, *2*(1), 24–31. https://doi.org/10.15294/ajpesh.v2i1.55227

Smits-Engelsman, B. C. M., Smit, E., Doe-Asinyo, R. X., Lawerteh, S. E., Aertssen, W., Ferguson, G., & Jelsma, D. L. (2021). Inter-rater reliability and test-retest reliability of the Performance and Fitness (PERF-FIT) test battery for children: a test for motor skill related fitness. *BMC Pediatrics*, *21*(1), 119. https://doi.org/10.1186/s12887-021-02589-0

Subhaktiyasa, P. G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka. *Journal of Education Research*, *5*(4), 5599–5609. https://doi.org/10.37985/jer.v5i4.1747

Yogamalar, I., & Samuel, A. A. (2018). Classification of Shared Values for Educational Research. *International Journal of Human Resource Studies*, *8*(2), 47. https://doi.org/10.5296/ijhrs.v8i2.12818