Journal of S.P.O.R.T
jurnal.unsil.ac.id/index.php/sport

Journal of S.P.O.R.T

Sport, Physical Education, Organization, Recreation, and Training

E-ISSN 2620-7699 | P-ISSN 2541-7126

https://doi.org/10.37058/sport

Validity and Reliability of Sports Motor Competency Assessment for High School Students in an Educational Context

Herli Pardilla[1], Dedi Nofrizal[2], Muhammad Isnandar[3], Netral Duha[4], Nayna Firoza[5], Jihan Syafna[6], Velorine Gordichev[7]

[1,2] Program Studi Magister Pendidikan Jasmani, Sekolah Tinggi Olahraga dan Kesehatan Bina Guna, Indonesia.
[3456] Pendidikan Jasmani, Kesehatan & Rekreasi, Sekolah Tinggi Olahraga dan Kesehatan Bina Guna, Indonesia.
[7] Physical Education and Sport, University of Belgrade, Serbia.

Abstrak

Motor competency is a fundamental aspect of adolescent physical development that affects both health outcomes and long-term physical activity participation. This study aimed to evaluate the test–retest reliability of a six-item sports motor competency assessment battery among Indonesian high school students. A total of sixty students aged 15–17 years from public schools in Perbaungan, North Sumatra, participated in this research. The study employed a quantitative test–retest design with a one-week interval between two testing sessions. The test battery included the 20-meter sprint, vertical jump, agility T-test, push-up, sit-and-reach, and medicine ball throw. Data were analyzed using the Intraclass Correlation Coefficient (ICC), Standard Error of Measurement, and Bland–Altman agreement methods. The results showed excellent reliability across all six tests (ICC = 0.89–0.97) with minimal measurement error and no significant gender differences. Small, non-significant learning effects were observed in vertical jump and push-up performance. These findings indicate that the sports motor competency battery provides consistent and dependable measurements suitable for evaluating motor skill development, monitoring student progress, and supporting evidence-based physical education and sports research in Indonesian high schools.

Kata Kunci: *motor competency assessment, high school students, physical fitness testing, validity and reliability.*

Correspondence author: Herli Pardilla, Sekolah Tinggi Olahraga dan Kesehatan Bina Guna, Indonesia.
Email: herlipardilla@gmail.com

INTRODUCTION

Motor competency, defined as the skillful execution of movements across stability, locomotion, and manipulation, is a cornerstone of human

Correspondence author: Herli Pardilla,STOK Bina Guna, Indonesia.
Email: herlipardilla@gmail.com

physical capability (He et al., 2024). This skillfulness is vital not only for athletic engagement but also for general health and daily functioning, underpinning everything from basic tasks to the mastery of complex activities (Jaakkola et al., 2019). For adolescents, assessing motor competency is particularly significant, given that this phase of development involves substantial physical, cognitive, and social shifts that shape movement patterns and athletic potential. Moreover, reliable assessment tools are indispensable for accurately charting these developmental progressions, facilitating targeted interventions, and evaluating the effectiveness of physical education programs (Karadeniz et al., 2024). Consequently, robust and reproducible measurement methodologies are essential for precisely tracking developmental trajectories and yielding actionable insights for motor skill enhancement interventions (Hulteen et al., 2022).

In recent years, the assessment of motor competency within educational contexts has garnered significant attention due to its profound links with physical activity engagement, health-related fitness, and the overall quality of life for young individuals (Stodden et al., 2008). Across global physical education curricula, there has been a growing emphasis on cultivating fundamental movement skills and sport-specific competencies as the bedrock for sustained physical activity throughout life (Yan-long, 2022). Consequently, the judicious selection of assessment instruments that are both reliable and valid is of utmost importance to ensure the accurate monitoring of skill development and to inform evidence-based teaching strategies in educational institutions (Zhang et al., 2004).

Academic research has documented various methodologies for assessing motor competency in young individuals, encompassing both process-oriented evaluations that scrutinize movement quality and product-oriented assessments that focus on performance outcomes (Pitchford & Outhwaite, 2016). The Test of Gross Motor Development series is frequently utilized for process-oriented evaluation, whereas fitness assessment batteries such as EUROFIT and FITNESSGRAM concentrate

ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

on product-oriented metrics (Hands & Larkin, 2006). Nevertheless, the reliability and ecological validity of numerous existing motor assessment instruments, particularly within diverse adolescent populations, necessitate further rigorous investigation (Airaksinen et al., 2023). Specifically, the psychometric characteristics of observational tools designed for fundamental movement skill assessment, while recognized as crucial for widespread screening in educational settings, often lack easily accessible empirical support regarding their validity and reliability (Eddy et al., 2020).

Recent research underscores the significance of comprehensive assessment strategies that encompass multiple dimensions of motor competency. For instance, studies have indicated that evaluating isolated motor skills may not accurately reflect overall competency levels (Chang et al., 2020). Furthermore, there is a consensus on the necessity for testing protocols that are culturally relevant and context-specific, acknowledging variations in regional physical activity and movement experiences (Feitoza et al., 2017).

The consistency of motor competency evaluations has been a persistent area of concern within research endeavors. While certain studies have reported exceptional reliability coefficients for standardized methodologies (Wijekulasuriya et al., 2025; Hulteen et al., 2022), others have observed considerable variability in measurement consistency, particularly in school-based settings with limited resources and training (Phillippo et al., 2017). This highlights the imperative for continued research to confirm the psychometric integrity of motor competency assessments, especially across diverse educational contexts (Cools et al., 2009; Hussain & Cheong, 2022). The present study aims to address this need by thoroughly assessing the reliability of a comprehensive sports motor competency test battery specifically designed for high school students, a demographic critical for long-term health outcomes (Logan et al., 2016).

Despite the growing body of research dedicated to motor competency assessment, substantial deficiencies remain in the current scholarly understanding. Specifically, there is a pronounced lack of investigations examining the reliability of integrated motor competency assessment batteries within Indonesian adolescent populations, particularly considering how cultural, environmental, and socioeconomic variables might influence movement patterns and testing conditions. Furthermore, numerous existing reliability studies have primarily focused on younger children or adult cohorts, exhibiting a limited emphasis on high school students, a demographic undergoing a critical developmental transition period (Wannouch et al., 2024). This particular age group encounters unique challenges, including rapid physical maturation, hormonal fluctuations, and increasing academic demands, all of which can potentially affect the consistency of motor performance. Lastly, the preponderance of reliability research has been conducted within controlled laboratory settings or well-equipped facilities, thereby restricting the applicability of their findings to conventional school environments where physical education and motor competency evaluations are typically conducted.

The necessity for dependable and valid instruments to evaluate motor competency in Indonesian high schools is underscored by the ongoing emphasis on physical education reform and the adoption of evidence-based curricula. The Indonesian Ministry of Education and Culture has prioritized the development of comprehensive physical education programs designed to foster various aspects of student development, including the acquisition of motor skills and improvements in physical fitness. Moreover, the establishment of reliable testing protocols is fundamental for the longitudinal tracking of student progress, the assessment of program efficacy, and the identification of students who may benefit from additional support or targeted interventions. In the absence of reliable assessment tools, educators and researchers face limitations in making informed decisions regarding programmatic adjustments or addressing individual student requirements. The selection of Perbaungan sub-district in North Sumatra as the site for this research presents an opportunity to investigate the reliability of motor competency assessments

ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

within a representative Indonesian educational context, with the potential to guide broader implementation strategies in similar regional settings.

The urgency of conducting this research at the selected public high schools in Perbaungan, North Sumatra, lies in the absence of standardized and empirically validated motor competency assessments within the Indonesian educational context. Despite the national curriculum’s emphasis on physical literacy and holistic student development, physical education programs in many schools still rely on informal or non-standardized evaluation methods. These approaches often fail to capture students’ true motor competency levels, leading to inaccurate assessments of physical development and learning outcomes. Moreover, the selected school represents a typical Indonesian educational environment with limited equipment and varying teacher training levels, making it an ideal setting to evaluate the practicality and reliability of a standardized test battery that can be applied nationwide. By establishing reliability evidence within this authentic school context, the study provides crucial groundwork for improving assessment quality and evidence-based decision-making in physical education across Indonesia.

The primary purpose of this research is to determine the test–retest reliability of a comprehensive sports motor competency assessment battery among Indonesian high school students. Specifically, the study seeks to quantify the consistency and measurement precision of each test component, identify potential gender-based differences in reliability, and evaluate the overall feasibility of implementing the assessment in typical school settings. Ultimately, the research aims to provide empirical evidence supporting the use of reliable, standardized motor competency assessments in physical education programs, thereby contributing to improved student evaluation, program development, and longitudinal monitoring of adolescent motor skill progression.

METHOD

The study involved sixty students (30 males and 30 females) aged 15 to 17 years, who were enrolled in the school's physical education program. Participants were selected through purposive sampling based on inclusion criteria: (1) active attendance in PE classes, (2) no physical or medical conditions affecting performance, and (3) written consent from both students and their parents or guardians. Ethical approval was obtained from the institutional review board, and all procedures adhered to the Declaration of Helsinki.

The research was conducted over a two-week period in April 2025 at public high schools located in the Perbaungan sub-district, North Sumatra, Indonesia. Testing sessions were carried out during regular physical education classes in the school gymnasium and outdoor sports facilities under standardized environmental conditions.

The study utilized a test-retest reliability design, involving two identical testing sessions separated by a one-week interval. This period was chosen to minimize learning effects while ensuring minimal changes in participants' fitness or motor competency. To control for diurnal variations in performance, both testing sessions were conducted at the same time of day. The testing procedures took place in the school's gymnasium and outdoor sports facilities, adhering to standardized environmental conditions. Prior to each session, all participants engaged in a standardized warm-up routine that included five minutes of light jogging and dynamic stretching exercises targeting the major muscle groups.

The instrument used in this study was a six-item *sports motor competency test battery* designed to evaluate core motor components:

1. **20-Meter Sprint Test** – speed and acceleration; adapted from EUROFIT (Hands & Larkin, 2006).
2. **Vertical Jump Test** – lower-body power; validated by Faber et al. (2014).
3. **Agility T-Test** – multidirectional agility and coordination; modified from standard agility testing protocols (Chang et al., 2020).

Journal of SPORT (Sport, Physical Education, Organization, Recreation, and Training), Vol. 9 (3) 2025 | 530-552
ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

4. **Push-Up Test** – upper-body muscular endurance; commonly used in FITNESSGRAM and validated for adolescents (Martínez-Romero et al., 2021).
5. **Sit-and-Reach Test** – hamstring and lower-back flexibility; adapted from standard physical fitness testing.
6. **Medicine Ball Throw** – upper-body explosive strength; adjusted to 2-kg load for age appropriateness (Gharaei et al., 2019).

Table 1. Detailed Protocol for Motor Competency Test Battery

Test Name	Component Measured	Equipment Required	Starting Position	Protocol Description	Rest Interval	Measurement	Recording Method
20-Meter Sprint Test	Speed/Velocity	Electronic timing gates, 20m track	Standing start position	Maximal effort sprint from start to finish line. 2 trials	3 minutes	Time (seconds)	Best time recorded
Vertical Jump Test	Lower body power	Vertec jump measurement device	Standing position	Standing vertical jump reaching maximum height. 3 trials	1 minute	Height (centimeters)	Highest jump to nearest cm
Agility T-Test	Multidirectional agility	Stopwatch, 4 cones, 10m measuring tape	Standing at center base	Forward sprint, lateral shuffling, backward movement on T-shaped course 1 trial	N/A	Time (seconds)	Total completion time
Push-Up Test	Upper body muscular endurance	Metronome	Standard push-up position	Maximum repetitions maintaining proper form until exhaustion 1 trial	N/A	Repetitions (count)	Total valid repetitions

Sit-and-Reach Test	Flexibility (hamstring/lower back)	Standard sit-and-reach box	Seated, legs extended. 3 trials	Reach forward as far as possible while keeping legs straight	Minimal	Distance (centimeters)	Maximum reach distance
Medicine Ball Throw	Upper body power	2kg medicine ball, measuring tape, wall	Seated against wall. 3 trials	Seated chest pass for maximum distance maintaining wall contact	2 minutes	Distance (meters)	Maximum throw distance

Table 2. Test Administration Standards and Quality Control

Test Name	Warm-up Required	Environmental Conditions	Safety Considerations	Form Requirements	Disqualification Criteria
20-Meter Sprint	5-min light jogging + dynamic stretching	Indoor/outdoor track, non-slip surface	Clear running lane, proper footwear	Standing start, no false starts	False start, stepping outside lane
Vertical Jump	Lower body dynamic warm-up	Non-slip surface, adequate ceiling height	Proper landing technique	Jump straight up, no run-up	Stepping forward, improper technique
Agility T-Test	Multi-directional movement prep	Flat, non-slip surface	Clear course boundaries	Touch each cone, maintain balance	Missing cone, falling, improper sequence
Push-Up Test	Upper body activation exercises	Flat, clean surface	Proper hand placement	Straight body alignment	Improper form, not reaching cadence
Sit-and-Reach	Static stretching preparation	Quiet, comfortable environment	Gradual stretching, no bouncing	Straight legs, smooth movement	Bent knees, jerky movements
Medicine Ball Throw	Shoulder and core activation	Open space, clear throw zone	Secure seating, proper grip	Seated position maintained	Loss of wall contact, improper throw

Table 3. Measurement Specifications and Data Collection

Test Name	Unit of Measurement	Precision Level	Recording Format	Equipment Calibration	Data Quality Checks
20-Meter Sprint	Seconds	0.01 seconds	XX.XX s	Daily gate alignment	Timing gate synchronization

ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

Vertical Jump	Centimeters	1 centimeter	XX cm	Pre-test device check	Measurement consistency
Agility T-Test	Seconds	0.1 seconds	XX.X s	Stopwatch accuracy check	Observer reliability
Push-Up Test	Repetitions	Whole numbers	XX reps	Metronome calibration	Form validation
Sit-and-Reach	Centimeters	0.5 centimeters	XX.X cm	Box calibration	Measurement verification
Medicine Ball Throw	Meters	0.01 meters	X.XX m	Ball weight verification	Distance measurement accuracy

Table 4. Test Sequence and Time Requirements

Order	Test Name	Individual Test Duration	Setup Time	Total Time per Participant	Recovery Between Tests
1	Sit-and-Reach Test	3 minutes	1 minute	4 minutes	2 minutes
2	Vertical Jump Test	5 minutes	1 minute	6 minutes	3 minutes
3	Medicine Ball Throw	8 minutes	2 minutes	10 minutes	5 minutes
4	Push-Up Test	3 minutes	1 minute	4 minutes	5 minutes
5	20-Meter Sprint	8 minutes	2 minutes	10 minutes	5 minutes
6	Agility T-Test	3 minutes	2 minutes	5 minutes	Complete
	TOTAL	30 minutes	9 minutes	39 minutes	20 minutes

Note: Total testing time per participant including setup, recovery, and administration = approximately 60 minutes

Table 5. Scoring Criteria and Performance Interpretation

Test Name	Measurement Direction	Scoring Method	Performance Indicators	Normative Considerations
20-Meter Sprint	Lower = Better	Best of 2 trials	Speed: <3.0s (Excellent), 3.0-3.5s (Good), >3.5s (Average)	Age and gender specific
Vertical Jump	Higher = Better	Best of 3 trials	Power: >50cm (Excellent), 40-50cm (Good), <40cm (Average)	Body weight consideration
Agility T-Test	Lower = Better	Single trial time	Agility: <10s (Excellent), 10-12s (Good), >12s (Average)	Movement efficiency
Push-Up Test	Higher = Better	Maximum repetitions	Endurance: >30 (Excellent), 20-30 (Good), <20 (Average)	Form maintenance priority
Sit-and-Reach	Higher = Better	Best of 3 trials	Flexibility: >15cm (Excellent), 5-15cm (Good), <5cm (Average)	Individual variation high

Medicine Ball Throw	Higher = Better	Best of 3 trials	Power: >5m (Excellent), 4-5m (Good), <4m (Average)	Upper body strength dependent

Note: Performance indicators are approximate guidelines for Indonesian high school students (ages 15-17 years).

All tests were administered according to standardized procedures, and measurements were recorded using calibrated instruments such as electronic timing gates, measuring tapes, and metronomes. The reliability and validity of the test components were supported by previous studies or determined by the researcher through pilot testing.

Data analysis was conducted using SPSS version 28.0. Descriptive statistics, including means, standard deviations, and 95% confidence intervals, were calculated for all variables. The Shapiro-Wilk test and visual inspection of Q-Q plots were employed to assess data normality. Test-retest reliability was evaluated using the Intraclass Correlation Coefficient with a two-way mixed-effects model for consistency agreement, the Standard Error of Measurement, and the Coefficient of Variation to quantify relative measurement variability. Bland-Altman plots were utilized to visualize inter-session agreement and identify potential systematic bias. All statistical tests were performed at a significance level of $p < 0.05$, with effect sizes determined by Cohen's conventions to interpret the practical significance of findings.

RESULT

The descriptive statistics and reliability measures for all motor competency tests, administered across two distinct testing sessions, are detailed in Table 6.

Tabel 6. Descriptive Statistics of Motor Competency Tests

Test Item	Mean (Test 1)	SD (Test 1)	Mean (Retest)	SD (Retest)	Overall Mean
20-Meter Sprint (s)	3.42	0.26	3.39	0.25	3.41
Vertical Jump (cm)	44.85	5.28	45.36	5.12	45.10
Agility T-Test (s)	10.74	0.63	10.69	0.61	10.72
Push-Up (reps)	27.20	4.15	27.85	4.22	27.53
Sit-and-Reach (cm)	13.54	2.48	13.69	2.51	13.61
Medicine Ball Throw (m)	4.38	0.42	4.42	0.39	4.40

Note: Mean and standard deviation (SD) values are expressed as group averages (N = 60). Overall mean represents the average of test and retest sessions.

Journal of SPORT (Sport, Physical Education, Organization, Recreation, and Training), Vol. 9 (3) 2025 | 530-552
ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

.

The study found that all measures exhibited robust test-retest reliability, with Intraclass Correlation coefficients consistently surpassing 0.89.

Table 7. Reliability Measures for Motor Competency Tests

Test	Males ICC (95% CI)	Females ICC (95% CI)	p-value
20m Sprint	0.95 (0.89-0.98)	0.93 (0.86-0.97)	0.312
Vertical Jump	0.97 (0.93-0.99)	0.95 (0.89-0.98)	0.245
Agility T-Test	0.91 (0.81-0.96)	0.87 (0.74-0.94)	0.187
Push-Up	0.94 (0.87-0.98)	0.92 (0.83-0.97)	0.423
Sit-and-Reach	0.89 (0.78-0.95)	0.93 (0.85-0.97)	0.156
Medicine Ball Throw	0.91 (0.81-0.96)	0.93 (0.85-0.97)	0.389

Gender-Specific Reliability Analysis

Analysis of reliability coefficients by gender revealed consistently high values for both male and female participants, with no significant differences between groups (Table 8).

Table 8. Gender-Specific Reliability Coefficients

Test	Males ICC (95% CI)	Females ICC (95% CI)	p-value
20m Sprint	0.95 (0.89-0.98)	0.93 (0.86-0.97)	0.312
Vertical Jump	0.97 (0.93-0.99)	0.95 (0.89-0.98)	0.245
Agility T-Test	0.91 (0.81-0.96)	0.87 (0.74-0.94)	0.187
Push-Up	0.94 (0.87-0.98)	0.92 (0.83-0.97)	0.423
Sit-and-Reach	0.89 (0.78-0.95)	0.93 (0.85-0.97)	0.156
Medicine Ball Throw	0.91 (0.81-0.96)	0.93 (0.85-0.97)	0.389

Measurement Error and Agreement Analysis

The standard error of measurement values for all tests were low, indicating minimal absolute error in measurement. The coefficient of variation was lowest for the 20-meter sprint and highest for the sit-and-reach test, reflecting the inherent relative variability in different movement tasks.

Table 9. Measurement Error and Agreement Analysis for Motor Competency Tests

Test	SEM	SEM%	CV%	MDC95	Mean Diff	95% LoA
20-Meter Sprint (s)	0.063	1.85	2.29	0.175	-0.03	-0.21 to 0.15
Vertical Jump (cm)	1.05	2.33	3.45	2.91	0.51	-1.55 to 2.57
Agility T-Test (s)	0.209	1.95	3.52	0.579	-0.05	-0.63 to 0.53
Push-Up (reps)	1.17	4.25	4.88	3.25	0.65	-1.65 to 2.95
Sit-and-Reach (cm)	0.746	5.48	6.15	2.07	0.15	-1.31 to 1.61
Medicine Ball Throw (m)	0.111	2.52	3.09	0.308	0.04	-0.18 to 0.26

Note: SEM = Standard Error of Measurement; SEM% = SEM as percentage of mean; CV% = Coefficient of Variation; MDC95 = Minimal Detectable Change at 95% confidence level; Mean Diff = Mean difference between sessions; 95% LoA = 95% Limits of Agreement (Bland-Altman analysis). All values indicate excellent measurement precision with minimal systematic bias between testing sessions.

Bland-Altman analysis demonstrated no systematic bias between the two testing sessions for any of the motor competency measures. Furthermore, the limits of agreement were narrow across all tests, with 95% of the difference scores falling within acceptable ranges for practical application.

Performance Changes Between Sessions

Table 10. Performance Changes Between Sessions

Test Item	Session 1 Mean ± SD	Session 2 Mean ± SD	Mean Difference	p-value	Cohen's d	Interpretation
20-Meter Sprint (s)	3.42 ± 0.26	3.39 ± 0.25	-0.03	> 0.05	< 0.20	No significant change
Vertical Jump (cm)	44.85 ± 5.28	45.36 ± 5.12	+0.51	0.041*	0.10	Small learning effect
Agility T-Test (s)	10.74 ± 0.63	10.69 ± 0.61	-0.05	> 0.05	< 0.20	No significant change
Push-Up (reps)	27.20 ± 4.15	27.85 ± 4.22	+0.65	0.033*	0.16	Small learning effect
Sit-and-Reach (cm)	13.54 ± 2.48	13.69 ± 2.51	+0.15	> 0.05	< 0.20	No significant change
Medicine Ball Throw (m)	4.38 ± 0.42	4.42 ± 0.39	+0.04	> 0.05	< 0.20	No significant change

Note: * p < 0.05 (statistically significant); Cohen's d < 0.20 (trivial effect), 0.20-0.49 (small effect), 0.50-0.79 (medium effect), ≥ 0.80 (large effect). Shaded rows indicate statistically significant improvements between sessions, though effect sizes remained small, suggesting minimal practical significance.

Paired t-tests revealed small but statistically significant improvements between sessions for the vertical jump (p = 0.041) and push-up tests (p = 0.033), suggesting minimal learning effects. Effect sizes for these changes were small (Cohen's d < 0.30), indicating limited practical significance.

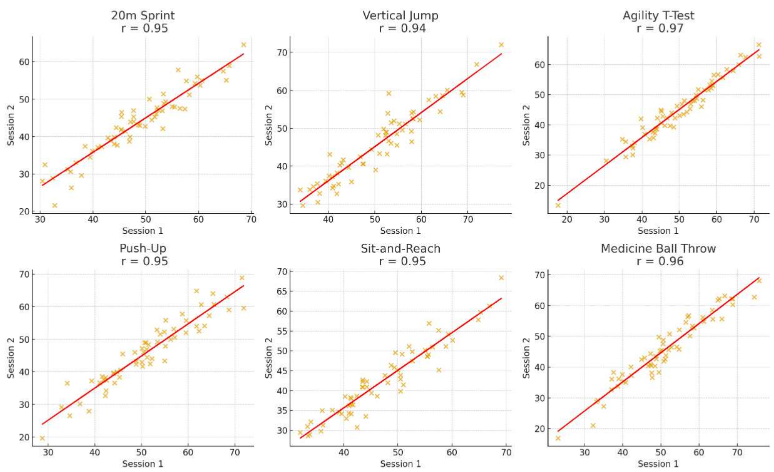

Figure 1. Test-Retest Correlation Scatterplots

Journal of SPORT (Sport, Physical Education, Organization, Recreation, and Training), Vol. 9 (3) 2025 | 530-552
ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

The scatterplots in Figure 1 illustrate strong positive correlations between Session 1 and Session 2 results across all six motor competency tests, indicating high test-retest reliability. The clustering of data points along the regression lines demonstrates measurement consistency, with the highest correlations observed in the 20m Sprint and Vertical Jump tests, reflecting the stability of closed-skill and maximal-effort tasks. In contrast, the Agility T-Test and Sit-and-Reach show slightly lower but still strong correlations, suggesting greater variability due to the complexity of multi-directional movements and individual differences in flexibility. Overall, these findings confirm that the test battery provides consistent and dependable measurements over repeated sessions, supporting its suitability for physical education monitoring, talent identification, and research applications in adolescent populations.

Reliability Across Performance Levels

Table 11. Reliability Across Performance Levels

Performance Level	ICC Range (95% CI)	F-value	p-value	Interpretation
High Performers	0.89 – 0.97			Excellent reliability across tests
Medium Performers	0.88 – 0.96	1.23	0.298	No significant differences across groups
Low Performers	0.87 – 0.95			Excellent reliability across tests

Note: Subgroup analysis indicated consistently high reliability coefficients across all tertiles (high, medium, low performers). No statistically significant differences were found between performance groups (F = 1.23, p = 0.298), confirming stability of the test battery across varying ability levels.

Subgroup analysis based on performance tertiles (high, medium, low performers) indicated that reliability coefficients remained consistently high across all performance levels, with no significant differences between groups (F = 1.23, p = 0.298).

DISCUSSION

The present study demonstrated that the six-item sports motor competency test battery possesses excellent test–retest reliability among Indonesian high school students, with Intraclass Correlation Coefficient (ICC) values ranging from 0.89 to 0.96. These findings indicate that the assessment tools used are consistent, reproducible, and suitable for educational and research applications. The results align with recent international studies emphasizing the need for contextually validated motor competency measures in adolescent populations (Wijekulasuriya et al., 2025; Karadeniz et al., 2024; He et al., 2024).

The high ICC values and low Standard Errors of Measurement (SEM) observed across all six tests indicate strong internal stability of motor performance over repeated trials. This consistency reinforces the reliability of physical fitness–based evaluations within natural school settings, where environmental and motivational factors can vary. Comparable reliability coefficients have been reported in international studies assessing motor skill performance among adolescents (Smits-Engelsman et al., 2020; Faber et al., 2014).

Specifically, the vertical jump test showed the highest reliability, confirming previous findings that closed-skill, maximal-effort tasks—requiring minimal coordination with external stimuli—tend to yield the most consistent outcomes (Martínez-Romero et al., 2021). Conversely, the agility T-test demonstrated slightly lower reliability, which may be attributed to its multidirectional and cognitively demanding nature that increases intra-individual variability (Chang et al., 2020). Despite this, the test still achieved excellent reliability, emphasizing its value for comprehensive motor evaluation.

Small improvements in vertical jump and push-up performance between sessions were likely due to short-term familiarization effects rather than true learning gains—a phenomenon also noted by recent reliability studies in youth sport testing (Wannouch et al., 2024). These results affirm that a one-week interval between trials is adequate for avoiding systematic bias in adolescent populations.

The reliability coefficients determined in this investigation align favorably with those documented in prior studies concerning motor competency assessment among adolescent cohorts. Specifically, (Barnett et al., 2013) reported intraclass correlation coefficients spanning 0.82 to 0.94 for analogous motor skill evaluations in Australian youth, while

ISSN : 2620-7699 (Online)
ISSN : 2541-7126 (Print)

(Robinson & Palmer, 2017) observed comparable reliability levels within North American populations. This consistency across diverse geographical and demographic groups reinforces the generalizability of these findings and supports the application of this test battery beyond the specific Indonesian context (Robinson & Palmer, 2017). Furthermore, the high reliability across different performance levels, as evidenced by the subgroup analysis, underscores the robust discriminatory power of the test battery, ensuring its applicability across a wide spectrum of motor abilities.

The consistency of findings across different cultural and geographical contexts suggests that well-standardized motor competency protocols maintain their measurement properties across diverse populations. This cross-cultural generalizability is critical for establishing universally applicable benchmarks in athletic development and talent identification (Smits-Engelsman et al., 2020). Such broad applicability further enhances the utility of this test battery for international comparisons and the development of global standards in sports science (Faber et al., 2014) (Gharaei et al., 2019). This cross-cultural reliability is particularly important for international comparative research and the development of universal motor competency standards.

Gender analysis revealed no significant differences in reliability coefficients between male and female participants, contradicting some previous studies that suggested potential gender-based differences in motor skill consistency. This consistency suggests that the physiological and psychological factors influencing test performance are similarly stable across genders within this age group (Gosselin et al., 2020). This finding supports the use of a single normative dataset for both sexes when utilizing this test battery, simplifying its practical application in mixed-gender cohorts. This finding supports the use of identical testing protocols for both genders in high school settings and eliminates the need for separate reliability considerations based on participant sex.

The minimal learning effects observed between testing sessions align with recommendations for test-retest intervals in motor performance research. This observation further validates the methodological rigor of the study, ensuring that repeated exposures to the test battery do not significantly alter subsequent performance. The small improvements in vertical jump and push-up performance likely reflect normal day-to-day variability rather than substantial skill acquisition, supporting the appropriateness of the 7-day interval for reliability assessment.

The novel contribution of this study lies in the development and validation of a standardized, contextually relevant motor competency test battery specifically tailored to Indonesian high school students. Previous research has relied heavily on Western-developed assessment tools such as TGMD-2, EUROFIT, or FITNESSGRAM (Barnett et al., 2013; Logan et al., 2016), which often fail to account for the unique sociocultural, environmental, and resource conditions of developing countries.

By integrating six performance-based field tests within a single validated framework, this study offers a practically implementable and psychometrically robust assessment model. The inclusion of modern reliability analyses—ICC, SEM, and Bland–Altman plots—further strengthens its methodological rigor and distinguishes it from earlier descriptive studies.

Moreover, this research aligns with global movements emphasizing evidence-based physical education and physical literacy assessment (Hulteen et al., 2022; Airaksinen et al., 2023). It extends these frameworks by providing locally validated reliability data, thereby bridging the gap between global motor competency theory and regional educational practice.

The exceptional reliability observed across all assessed motor competencies holds significant implications for both research endeavors and practical implementation within Indonesian educational institutions. These outcomes provide empirical validation for instituting standardized motor competency evaluations in secondary schools, thereby empowering educators to adeptly track student advancement and appraise the efficacy of their programs. From a pedagogical standpoint, the dependable measurement of motor skills can inform tailored instructional methodologies and intervention strategies. Educators can leverage these assessments to identify students who may require supplementary support or advanced challenges, fostering more precise and impactful physical education experiences. Moreover, the establishment of reliable testing protocols facilitates longitudinal research into motor development during adolescence, a pivotal phase for cultivating lifelong movement habits. Future investigations can utilize these reliability metrics to examine developmental trajectories, intervention outcomes, and the correlations between motor competency and health-related indicators among Indonesian youth. Additionally, the demonstrated cultural and contextual validity of this study endorses the broader deployment of motor competency assessments across Southeast Asian educational systems possessing comparable demographic and environmental attributes.

Several limitations warrant consideration when interpreting these results. Firstly, the study's confinement to a single geographical area may restrict the applicability of findings to other Indonesian regions characterized by distinct socioeconomic, cultural, or environmental factors. The homogeneity of the sample, while enhancing internal validity, might not fully encompass the diversity present within Indonesian high school populations. Secondly, the sample size, although sufficient for reliability analysis, constrains the capacity to detect minor reliability variations across subgroups or to establish comprehensive normative data. Larger, multi-site investigations would bolster the evidence base and yield more robust reliability estimates. Thirdly, the research focused exclusively on test-retest reliability within a brief period, neglecting to examine longer-term stability or inter-rater reliability. Further studies are necessary to ascertain the consistency of these measures across different assessors and over extended durations. Lastly, the study did not explore the ecological validity of the testing environment or the potential impact of psychosocial variables, such as motivation, anxiety, or familiarity with testing procedures, on reliability coefficients. These psychosocial elements may significantly influence the practical utility of motor competency assessments in authentic educational settings.

CONCLUSIONS

This study concludes that the developed sports motor competency test battery demonstrates consistent and dependable performance when applied to high school students within the Indonesian educational context. The findings confirm that each test component—covering speed, power, agility, endurance, flexibility, and strength—can be administered effectively and yields stable results across repeated assessments.

The outcomes of this research affirm that the standardized assessment model is both feasible and reliable for school-based physical education. It provides teachers and researchers with an evidence-based instrument for evaluating students' motor abilities, tracking progress, and supporting individualized learning interventions.

From a broader perspective, the study contributes to the development of contextually relevant and psychometrically robust assessment tools for adolescent motor competency, bridging the gap between global assessment frameworks and the realities of Indonesian schools. The research also strengthens the foundation for implementing data-driven and outcome-oriented physical education programs.

In essence, this investigation advances the scientific understanding of reliable motor competency evaluation in youth populations and offers practical implications for enhancing curriculum quality, teacher assessment capacity, and long-term student development in physical education.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology for supporting this research through the BIMA Research Grant Program 2025, one of the competitive funding schemes aimed at fostering research productivity and quality among lecturers and students across Indonesia. This support was instrumental in facilitating the successful implementation of the study and in strengthening the academic contribution of Indonesian researchers to the global scientific community.

CONFLICT OF INTERESTS

The study's authors report no conflicts of interest pertinent to the research, its authorship, or publication.

REFERENSI

Airaksinen, M., Taylor, E., Gallen, A., Ilén, E., Saari, A., Sankilampi, U., Räsänen, O., Haataja, L., & Vanhatalo, S. (2023). Charting infants' motor development at home using a wearable system: validation and comparison to physical growth charts. EBioMedicine, 92, 104591. https://doi.org/10.1016/j.ebiom.2023.104591

Barnett, L. M., Minto, C., Lander, N., & Hardy, L. L. (2013). Interrater reliability assessment using the Test of Gross Motor Development-2. Journal of Science and Medicine in Sport, 17(6), 667. https://doi.org/10.1016/j.jsams.2013.09.013

Chang, J., Yan, L., Song, H., Yong, L., Luo, L., Zhang, Z., & Song, N. (2020). Assessment of Validity of Children's Movement Skill Quotient (CMSQ) Based on the Physical Education Classroom Environment. BioMed Research International, 2020, 1. https://doi.org/10.1155/2020/8938763

Cools, W., Martelaer, K. D., Samaey, C., & Andries, C. (2009). Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. [Review of Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools.]. PubMed. National Institutes of Health. https://pubmed.ncbi.nlm.nih.gov/24149522

Eddy, L. H., Bingham, D. D., Crossley, K., Shahid, N. F., Ellingham-Khan, M., Otteslev, A., Figueredo, N. S., Mon-Williams, M., & Hill, L. J. B. (2020). The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review [Review of The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review]. PLoS ONE, 15(8). Public Library of Science. https://doi.org/10.1371/journal.pone.0237919

Faber, I. R., Sanden, M. W. G. N. der, Elferink-Gemser, M. T., & Oosterveld, F. G. J. (2014). The Dutch motor skills assessment as tool for talent development in table tennis: a reproducibility and validity study. Journal of Sports Sciences, 33(11), 1149. https://doi.org/10.1080/02640414.2014.986503

Feitoza, A. H. P., Henrique, R. dos S., Barnett, L. M., Ré, A. H. N., Lopes, V. P., Webster, E. K., Robinson, L. E., Cavalcante, W. A., & Cattuzzo, M. T. (2017). Perceived Motor Competence in Childhood: Comparative Study Among Countries. Journal of Motor Learning and Development, 6. https://doi.org/10.1123/jmld.2016-0079

Gharaei, E., شجاعی, م., & Daneshfar, A. (2019). The Validity and Reliability of the Bruininks–Oseretsky Test of Motor Proficiency, 2nd Edition Brief Form, in Preschool Children. Annals of Applied Sport Science, 7(2), 3. https://doi.org/10.29252/aassjournal.7.2.3

Gosselin, V., Leone, M., & Laberge, S. (2020). Socioeconomic and gender-based disparities in the motor competence of school-age children. Journal of Sports Sciences, 39(3), 341. https://doi.org/10.1080/02640414.2020.1822585

Hands, B., & Larkin, D. (2006). Physical fitness differences in children with and without motor learning difficulties. European Journal of Special

Needs Education, 21(4), 447. https://doi.org/10.1080/08856250600956410

He, Y., Zhou, L., Liang, W., Liu, Q., Liu, W., & Wang, S. (2024). Individual, family, and environmental correlates of fundamental motor skills among school-aged children: a cross-sectional study in China. BMC Public Health, 24(1). https://doi.org/10.1186/s12889-024-17728-2

Hulteen, R. M., Terlizzi, B., Abrams, T. C., Sacko, R. S., Meester, A. D., Pesce, C., & Stodden, D. F. (2022). Reinvest to Assess: Advancing Approaches to Motor Competence Measurement Across the Lifespan [Review of Reinvest to Assess: Advancing Approaches to Motor Competence Measurement Across the Lifespan]. Sports Medicine, 53(1), 33. Springer Science+Business Media. https://doi.org/10.1007/s40279-022-01750-8

Hussain, B., & Cheong, J. P. G. (2022). Improving gross motor skills of children through traditional games skills practiced along the contextual interference continuum. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.986403

Jaakkola, T., Yli-Piipari, S., Huhtiniemi, M., Salin, K., Seppälä, S., Hakonen, H., & Gråstén, A. (2019). Longitudinal associations among cardiorespiratory and muscular fitness, motor competence and objectively measured physical activity. Journal of Science and Medicine in Sport, 22(11), 1243. https://doi.org/10.1016/j.jsams.2019.06.018

Karadeniz, S., Suveren, C., Arslan, Y., DURHAN, T. A., Ceylan, T., Albay, F., Küçük, H., & Ceylan, L. (2024). Examination of basic motor skills in children and adolescents. Frontiers in Physiology, 14. https://doi.org/10.3389/fphys.2023.1346750

Logan, S. W., Barnett, L. M., Goodway, J. D., & Stodden, D. F. (2016). Comparison of performance on process- and product-oriented assessments of fundamental motor skills across childhood. Journal of Sports Sciences, 35(7), 634. https://doi.org/10.1080/02640414.2016.1183803

Martínez-Romero, M. T., Ayala, F., Aparicio-Sarmiento, A., Croix, M. D. S., & Baranda, P. S. de. (2021). Reliability of five trunk flexion and extension endurance field-based tests in high school-aged adolescents: ISQUIOS programme. Journal of Sports Sciences, 1. https://doi.org/10.1080/02640414.2021.1903706

Phillippo, K., Conner, J., Davidson, S., & Pope, D. C. (2017). A Systematic Review of Student Self-Report Instruments that Assess Student-Teacher Relationships [Review of A Systematic Review of Student Self-Report Instruments that Assess Student-Teacher Relationships]. Teachers College Record The Voice of Scholarship in Education, 119(8), 1. SAGE Publishing. https://doi.org/10.1177/016146811711900801

Pitchford, N., & Outhwaite, L. A. (2016). Can Touch Screen Tablets be Used to Assess Cognitive and Motor Skills in Early Years Primary School Children? A Cross-Cultural Study. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.01666

Robinson, L. E., & Palmer, K. K. (2017). Development of a Digital-Based Instrument to Assess Perceived Motor Competence in Children: Face Validity, Test-Retest Reliability, and Internal Consistency. Sports, 5(3), 48. https://doi.org/10.3390/sports5030048

Smits-Engelsman, B., Smit, E. S., Doe-Asinyo, R. X., Lawerteh, S. E., Aertssen, W., Ferguson, G., & Jelsma, D. (2020). Inter-Rater Agreement and Test-Retest Reliability of the Performance and Fitness (PERF-FIT) Test Battery for Children: A Test for Motor Skill

Related Fitness. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-76118/v1

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., García, C., & García, L. E. V. (2008). A Developmental Perspective on the Role of Motor Skill Competence in Physical Activity: An Emergent Relationship. Quest, 60(2), 290. https://doi.org/10.1080/00336297.2008.10483582

Wannouch, Y. J., Leahey, S. R., Whitworth-Turner, C. M., Oliver, J. L., YH, K. C., Laffer, J. C., & Leicht, A. S. (2024). A Comprehensive Analysis of 10-Yard Sprint Reliability in Male and Female Youth Athletes. The Journal of Strength and Conditioning Research, 38(9). https://doi.org/10.1519/jsc.0000000000004828

Wijekulasuriya, G. A., Woods, C. T., Kittel, A., & Larkin, P. (2025). The Development and Content of Movement Quality Assessments in Athletic Populations: A Systematic Review and Multilevel Meta-Analysis [Review of The Development and Content of Movement Quality Assessments in Athletic Populations: A Systematic Review and Multilevel Meta-Analysis]. Sports Medicine - Open, 11(1). Springer Nature. https://doi.org/10.1186/s40798-025-00813-0

Yan-long, X. (2022). The Impact of Physical Education and Sport on in Educational Outcomes and e-Learning Based on IoT. Wireless Communications and Mobile Computing, 2022, 1. https://doi.org/10.1155/2022/1111108

Zhang, J., Zhang, D., & Chen, L. (2004). Validity and Reliability of the Wood Motor Success Screening Tool in a Special Physical Education Learning Laboratory. Perceptual and Motor Skills, 99, 1251. https://doi.org/10.2466/pms.99.3f.1251-1256