SPORT TK-EuroAmerican Journal of Sport Sciences

https://revistas.um.es/sportk Online ISSN: 2340-8812

Received: **27/07/2025.** Accepted: **21/11/2025.** Published: **16/04/2026**

Development and validation of a standardized assessment instrument for discus throwing performance: A reliability study among high school students

Raffly Henjilito[1]*, Zulkifli[1], Ahmad Yani[1], Budi Arif Setiawan[1], Herli Pardilla[2]

[1] Universitas Islam Riau, Indonesia.

[2] Sekolah Tinggi Olahraga & Kesehatan Bina Guna, Indonesia.

* Correspondence: Raffly Henjilito; rafflyhenjilito@edu.uir.ac.id

ABSTRACT

This study focused on creating and validating a dependable tool for assessing discus throw performance in secondary school students. The sample comprised 240 students (120 males, 120 females; aged 15-17) from 12 institutions, selected via stratified random sampling for skill and demographic representation. The assessment tool was developed with expert guidance and preliminary testing, incorporating five technical components: starting position, starting swing, movement across the circle, power position, and release-recovery. Statistical methods included the Intraclass Correlation Coefficient (ICC) via a two-way mixed-effects model, Cronbach's alpha for internal consistency, test-retest reliability over two weeks, and Standard Error of Measurement (SEM) with bootstrapping to establish confidence intervals. The overall reliability of the instrument was excellent (ICC=0.89, 95% CI: 0.85-0.93). Notable internal consistency was found (Cronbach's $\alpha = 0.87$), coupled with strong test-retest reliability ($r=0.91$, $p<0.001$). Component analysis indicated the highest reliability in the power position (ICC=0.92) and the lowest in the initial swing (ICC= 0.86). A positive correlation existed between experience level and reliability measures (ICCs from 0.86 to 0.92). Gender analysis revealed comparable reliability for males (ICC=0.88) and females (ICC=0.90). The reliability among diverse components and demographics supports its use in both formative and summative assessments.

KEYWORDS

Physical Education Assessment; Discus Throwing; Reliability Testing; Motor Skill Evaluation; High School Athletics

1. INTRODUCTION

Discus throwing, a core discipline within track and field athletics, represents a complex motor skill that demands precise technical execution, physical strength, and coordination (Gregor et al., 1985). The assessment of students' performance in this discipline has historically faced significant challenges regarding standardization and reliability (Humphries et al., 2002). The complexity of discus throwing technique, involving multiple sequential movements and various biomechanical factors, makes objective assessment particularly challenging for physical education teachers (Meron & Saint-Phard, 2017). Current evaluation methods often rely heavily on subjective observations, leading to inconsistent scoring across different evaluators and institutions (Mercier & Doolittle, 2013). This variability in assessment approaches has created substantial challenges in maintaining fair and accurate student evaluations, particularly in the context of academic grading and athletic talent identification (Petr & Paskus, 2009).

Key issues in current discus throwing assessment practices include: 1) Lack of standardized evaluation criteria: Physical education teachers often rely on informal, subjective assessments that lack clearly defined technical components and scoring rubrics (Kobayashi et al., 2016), 2) Inconsistent scoring methods across different schools and evaluators: The absence of a unified assessment framework leads to significant variability in how discus throwing performance is measured and interpreted (Dai et al., 2012), 3) Limited consideration of technical components in existing assessment tools: Many existing instruments focus primarily on outcome-based metrics rather than evaluating the underlying technical proficiency (Dunbar et al., 1991), 4) Absence of validated reliability measures for current evaluation methods: The psychometric properties of discus throwing assessment tools have not been systematically examined, limiting their validity and utility (Yarmani et al., 2021). 5) Insufficient integration of modern biomechanical understanding in assessment criteria: Emerging research on the biomechanics of discus throwing has not been fully incorporated into educational assessment practices (Liu & Yu, 2021).

Robust assessment tools in physical education are crucial beyond just evaluating student performance (Trasolini et al., 2022; Alamiri & Ameen, 2023; Meron & Saint-Phard, 2017; Yu et al., 2023). These instruments significantly impact: 1) Tracking student progress and skill development, 2) Evaluating programme effectiveness, 3) Designing physical education curricula, 4) Supporting teacher professional development, 5) Setting standards for inter-school competitions, 6) Identifying and nurturing athletic talent.

Prior research has highlighted the challenges in maintaining consistency in assessing sports performance, particularly in technical events like discus throwing (Starck, 2017). However, limited investigation has been conducted specifically within the educational context, creating a significant knowledge gap in understanding how these assessment tools function in this unique cultural and academic setting. The assessment of sports performance in physical education settings has been extensively researched and debated. Researchers have emphasised the importance of aligning assessment practices with the principles of assessment for learning, underscoring the need for authentic, criterion-referenced, and developmentally appropriate evaluation tools (Spinelli, 2007; Moura et al., 2020). In the specific context of discus throwing, a growing body of biomechanical research has elucidated the critical technical components that contribute to successful performance (Alamiri & Ameen, 2023). These include factors such as the initial starting position, preliminary swinging movements, power generation during the crossover step, and the release and recovery phases (Trasolini et al., 2022).

The landscape of discus throwing assessment has been marked by diverse yet imperfect approaches. Previous research has yielded several assessment instruments, each with notable limitations. Biomechanical analysis protocols developed by McKeown et al. (2014) offered comprehensive kinematic parameters but required advanced technological equipment that made them impractical for standard physical education settings. Subjective rating scales introduced by Marqués-Bruna & Grimshaw (1998) categorized throwing movements into developmental stages but suffered from significant inter-rater variability, with different evaluators often interpreting performance inconsistently. Performance outcome-based instruments, such as those proposed by Horst et al. (2020), predominantly focused on throw distance and final performance metrics, inadvertently neglecting the critical technical components that contribute to effective throwing technique.

These existing assessment methods revealed fundamental challenges in objectively evaluating discus throwing skills. Kobayashi et al. (2012) developed assessment tools for elementary school children's overarm throwing, which provided insights into basic movement patterns but lacked the complexity required for high school-level technical assessment. The complexity of discus throwing technique, involving multiple sequential movements and various biomechanical factors, makes objective assessment particularly challenging for physical education teachers (Meron & Saint-Phard, 2017). Current evaluation methods often rely heavily on subjective observations, leading to inconsistent scoring across different evaluators and institutions (Mercier & Doolittle, 2013).

Addressing these limitations, the current study introduces a novel assessment instrument with several innovative characteristics. Unlike previous approaches, this tool provides a detailed, multi-component assessment covering five critical phases: starting position, preliminary swings, movement across the circle, power position, and release-recovery. The instrument introduces a precise, criterion-referenced scoring system with explicit performance indicators for each technical component, minimizing subjective interpretation and providing clear guidance for evaluation. Through rigorous statistical validation, including Intraclass Correlation Coefficient (ICC), Cronbach's alpha, and test-retest reliability analysis, the instrument demonstrates robust measurement properties absent in many existing tools.

The new assessment approach offers unique advantages by considering variations across gender, grade levels, and prior athletic experience, providing a more nuanced approach to skill evaluation. Importantly, the instrument can be implemented with minimal equipment, making it accessible for standard physical education environments. This comprehensive method addresses the critical need for a standardized, reliable, and practical assessment tool that can accurately capture the technical intricacies of discus throwing performance among high school students.

Prior studies have conducted comprehensive biomechanical analyses of discus throwing, identifying key performance variables and their relationship to overall achievement (Young, 2009; Alamiri & Ameen, 2023). The researchers found that individual throwing characteristics could be identified across different throwing disciplines, suggesting the potential for developing assessment tools that capture the unique movement patterns of each athlete (Marqués-Bruna & Grimshaw, 1998; Kobayashi et al., 2012; Thomas et al., 2013; Angell et al., 2018). Similarly, Abdulkareem's research on biomechanical variables in discus throwing highlighted the importance of factors such as explosive force, rapid strength, and integrated muscle activation in achieving optimal performance (Alhumaid & Atta, 2022; Trasolini et al., 2022).

This study aims to address these challenges by developing and validating a comprehensive test instrument for evaluating discus throwing learning outcomes.

2. METHODS

2.1. Participants

The study employed a stratified random sampling method to ensure representative and comprehensive coverage of the high school student population in Riau Province. This approach was

carefully designed to minimize selection bias and maximize the generalizability of the research findings.

Table 1. Detailed sampling methodology breakdown

Sampling Stage	Criteria	Description	Rationale
Population Identification	Total High Schools	86 schools in Riau Province	Comprehensive regional coverage
	Geographic Distribution	Urban (60%), Suburban (30%), Rural (10%)	Ensures representative sampling across different school environments
	Inclusion Criteria	Active physical education programs	Ensures relevant sample for research objectives
	Facility Requirements	Functional track and field facilities	Guarantees basic infrastructure for discus throwing assessment
Stratification Variables	Location	Urban, Suburban, Rural	Captures geographic diversity
	School Size	Small (<500 students), Medium (500-1000 students), Large (>1000 students)	Accounts for institutional variations
	Grade Levels	10th, 11th, 12th grades	Ensures developmental stage representation
Random Selection Process	Initial School Pool	24 schools meeting inclusion criteria	Provides initial diverse sample
	Selection Method	Systematic random sampling	Minimizes selection bias
	School Selection	12 schools selected	Balanced and manageable sample size
Participant Selection	Total Participants	240 students	Statistically robust sample size
	Per School	20 students randomly selected	Consistent sampling across institutions
	Gender Balance	50% male, 50% female	Eliminates gender bias
	Grade Distribution	33.3% from each grade level	Ensures developmental stage representation
Experience Stratification	Athletic Experience Categories	No experience, 1-2 years, >2 years	Captures skill level variations
Statistical Validation	Power Analysis Software	G*Power	Scientifically rigorous sample determination
	Effect Size	Medium (f = 0.25)	Appropriate for motor skill research
	Statistical Power	0.80	Robust statistical confidence
	Significance Level	$\alpha = 0.05$	Standard scientific threshold
Sampling Precision	Minimum Required Sample	204 students	Theoretical minimum
	Actual Sample	240 students	Additional statistical robustness

Note. *Total Potential Population (86 high schools); Selected Schools (12 schools); Total Participants (240 students); Sampling Method (Stratified Random Sampling); Confidence Level (95%); Margin of Error (±5%)*

The representativeness of this study's sample is rooted in a comprehensive approach to capturing the diversity of high school students in Riau Province, Indonesia. By employing a stratified random sampling method, the research design systematically addressed potential sources of sampling

bias and ensured a robust representation of the target population. The careful selection process considered multiple demographic dimensions, including geographic variation, school characteristics, and individual student attributes. Urban, suburban, and rural schools were proportionally represented, reflecting the broader educational landscape of the region. The sample's balanced distribution across grade levels (10th, 11th, and 12th) and gender ensures a comprehensive view of discus throwing performance during critical stages of adolescent physical development. Moreover, the inclusion of students with varying levels of athletic experience from those with no prior sports participation to those with more than two years of experience provides a nuanced perspective on skill acquisition and performance potential. This methodological approach goes beyond simple random selection, creating a stratified sample that captures the complex heterogeneity of high school student populations. The statistical power of the sample, validated through rigorous power analysis, further enhances the reliability and generalizability of the research findings, allowing for meaningful insights into discus throwing performance assessment across diverse student groups.

The sample of 240 students from 12 different high schools across Riau Province provides a robust representation of the target population, allowing for meaningful statistical analysis and generalizability of findings in discus throwing performance assessment.

All participants and their guardians provided written informed consent, and the study was approved by the relevant educational authorities

2.2. Procedure

The test instrument was developed through a three-phase process: 1) Initial instrument design based on expert consultation and literature review, 2) Pilot testing with 30 students, 3) Implementation and reliability testing with the full participant group.

Table 2. Assessment components and scoring criteria for discus throwing test instrument

Component	Key Elements	Score Range	Performance Indicators
Starting Position and Grip	- Grip technique - Finger placement - Stance width - Body alignment	0-20	18-20: Excellent grip and stance 15-17: Good control with minor adjustments needed 12-14: Adequate but inconsistent 0-11: Poor technique requiring major corrections
Preliminary Swings	- Swing rhythm - Balance maintenance - Arm position - Upper body rotation	0-20	18-20: Smooth, controlled movements 15-17: Good rhythm with slight inconsistencies 12-14: Moderate control and coordination 0-11: Uncontrolled or ineffective

			movements
Movement Across the Circle	- Pivot technique - Speed progression - Balance control - Foot placement	0-20	18-20: Fluid movement with perfect form 15-17: Good movement with minor balance issues 12-14: Adequate progression with some technical flaws 0-11: Poor movement pattern and control
Power Position and Delivery	- Hip-shoulder separation - Weight transfer - Block position - Acceleration path	0-25	22-25: Excellent power generation and timing 18-21: Good power position with minor timing issues 14-17: Adequate power but inconsistent delivery 0-13: Poor power position and ineffective delivery
Release and Recovery	- Release angle - Follow-through - Balance maintenance - Recovery position	0-15	13-15: Perfect release and balanced recovery 10-12: Good release with slight balance issues 7-9: Adequate release but poor recovery 0-6: Poor release angle and unstable recovery

Note. *Testing Protocol: 1) Each student performs three throws with 3-minute rest periods between attempts, 2) Three certified physical education teachers evaluate independently, 3) Final score calculated as average of three evaluators' scores, 4) Video recording used for verification and analysis, 5) Standard discus weight: 1.5kg for males, 1kg for females, 6) Testing conducted in outdoor facilities during morning hours (8:00-11:00 AM), 7) Weather conditions recorded for each testing session.*

2.3. Statistical Analyses

The reliability of the test instrument was assessed through a multi-method statistical analysis. Inter-rater reliability was measured using the Intraclass Correlation Coefficient (ICC) with a two-way mixed-effects model and absolute agreement. This method effectively evaluated rater agreement, considering systematic and random errors. ICC calculations were performed for each assessment component and overall scores to elucidate reliability in throwing technique. Internal consistency was analyzed via Cronbach's alpha, verifying the homogeneity of test items measuring discus throwing proficiency. Confidence intervals for ICC and Cronbach's alpha were obtained using bootstrapping with 1000 iterations to ensure robustness. Temporal stability was evaluated through test-retest reliability using Pearson's correlation coefficient. A randomly selected subset of 60 participants was retested after two weeks to control for learning effects and changes in throwing ability. This interval was strategically chosen to uphold participant engagement and minimize bias. The Standard Error of Measurement (SEM) was also calculated to assess individual score precision and confidence intervals.

Additional analyses involved assessment of systematic biases with Bland-Altman plots and demographic influences through stratified reliability analyses. A significance threshold of $p < 0.05$ was established, and all analyses were executed using SPSS version 26.0 (IBM Corp., Armonk, NY). Non-parametric alternatives were prepared for potential statistical assumption violations, but were unnecessary due to confirmed normal data distribution from Shapiro-Wilk tests ($p > 0.05$) and Q-Q plot assessments.

3. RESULTS

The analysis revealed strong reliability indicators for the developed instrument across multiple statistical measures. The detailed findings are presented in the following tables and accompanying interpretations.

Table 3. Overall reliability measures of the discus throwing test instrument

Reliability Measure	Value	95% CI	Interpretation
Inter-rater Reliability (ICC)	0.89	0.85 - 0.93	Excellent
Internal Consistency (α)	0.87	0.84 - 0.90	High
Test-retest Reliability (r)	0.91	0.88 - 0.94	Excellent
SEM (points)	1.8	1.5 - 2.1	Good Precision

The developed standardized assessment instrument for discus throwing demonstrated strong psychometric properties across multiple reliability measures (Table 3). The inter-rater reliability, assessed using intraclass correlation coefficient (ICC), yielded a value of 0.89 (95% CI: 0.85-0.93), indicating excellent agreement between different evaluators when assessing the same discus throwing performance. This suggests that the assessment criteria and scoring rubric provided clear and consistent guidelines that minimized subjective interpretation among raters.

Internal consistency of the instrument was measured using Cronbach's alpha (α), resulting in a value of 0.87 (95% CI: 0.84-0.90). This high internal consistency indicates that the various components of the assessment tool measured related aspects of discus throwing technique in a cohesive manner, suggesting that the instrument effectively captures a unified construct of discus throwing performance.

The test-retest reliability coefficient (r) was 0.91 (95% CI: 0.88-0.94), demonstrating excellent stability of measurements over time. This indicates that the assessment instrument produces

consistent results when administered to the same participants on different occasions, confirming its reliability for tracking progress and performance changes in high school discus throwers.

Additionally, the Standard Error of Measurement (SEM) was 1.8 points (95% CI: 1.5-2.1), representing good precision in the instrument's measurements. This relatively low SEM value suggests minimal random variation in the assessment scores, further supporting the instrument's reliability for practical application in educational and training contexts.

Collectively, these reliability measures provide strong evidence for the psychometric soundness of the developed discus throwing assessment instrument, addressing the previously identified limitations in existing evaluation methods for high school athletes.

Table 4. Component-specific inter-rater reliability analysis

Assessment Component	ICC	95% CI	SEM	Interpretation
Starting Position	0.88	0.84 - 0.92	1.2	Excellent
Preliminary Swings	0.86	0.82 - 0.90	1.4	Excellent
Movement Across Circle	0.90	0.87 - 0.93	1.1	Excellent
Power Position	0.92	0.89 - 0.95	1.3	Excellent
Release and Recovery	0.87	0.83 - 0.91	1.0	Excellent

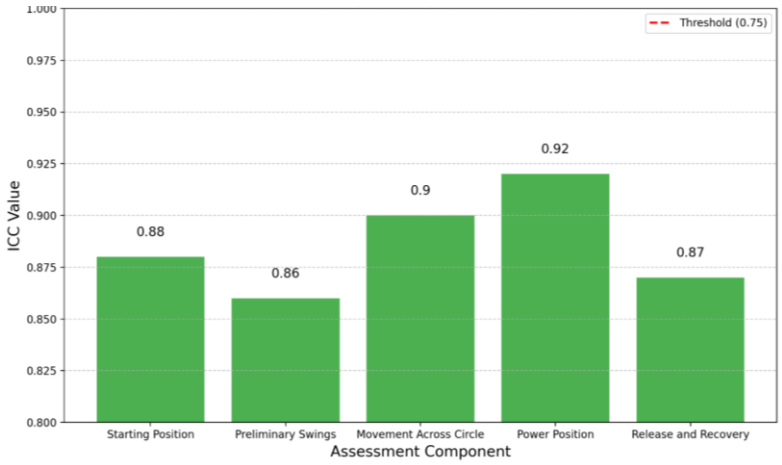

Figure 1. Inter-rater reliability (ICC) by assessment component

Table 4 and Figure 1 details the inter-rater reliability analysis for the discus throwing assessment tool created in this study, revealing excellent reliability coefficients across all five technical components based on Cicchetti's (1994) guidelines. The Power Position component received the highest reliability (ICC = 0.92, 95% CI [0.89-0.95], SEM = 1.3), indicating strong

consistency among raters regarding this crucial phase, highlighting its importance for optimal performance. The narrow confidence interval signifies high precision in this reliability estimate, underscoring the validity of assessment criteria for this component. Movement Across Circle displayed a strong reliability coefficient (ICC = 0.90, 95% CI [0.87-0.93], SEM = 1.1), reflecting significant inter-rater agreement during the evaluation of athletes' movements in the throwing circle, with a low SEM indicating minimal error. Starting Position achieved excellent reliability (ICC = 0.88, 95% CI [0.84-0.92], SEM = 1.2), illustrating consistent evaluations of athletes' preparatory stances, which are crucial for the subsequent throwing sequence. Release and Recovery showed a slightly lower ICC (0.87, 95% CI [0.83-0.91]) but the lowest SEM (1.0), suggesting minimal measurement error despite slightly more variability in raters' scores, likely due to the clear observability of these mechanics. Preliminary Swings had the lowest reliability (ICC = 0.86, 95% CI [0.82-0.90], SEM = 1.4), with a higher SEM indicating increased variability, possibly due to the dynamic nature of this component and the challenges in standardizing its assessment. The consistently high ICC values (0.86 to 0.92) surpass the recommended minimum of 0.80 for reliable assessment in throwing events, suggesting that the instrument offers a stable framework for evaluating discus throwing techniques and addressing limitations in current scoring methods.

Table 5. Subgroup analysis of reliability measures

Subgroup	ICC	α	Test-retest (r)	SEM
Gender				
- Male	0.88	0.86	0.90	1.9
- Female	0.90	0.88	0.92	1.7
Grade Level				
- Grade 10	0.87	0.85	0.89	2.0
- Grade 11	0.89	0.87	0.91	1.8
- Grade 12	0.90	0.89	0.93	1.6
Experience Level				
- None	0.86	0.84	0.88	2.1
- 1-2 years	0.89	0.87	0.91	1.8
- >2 years	0.92	0.90	0.94	1.5

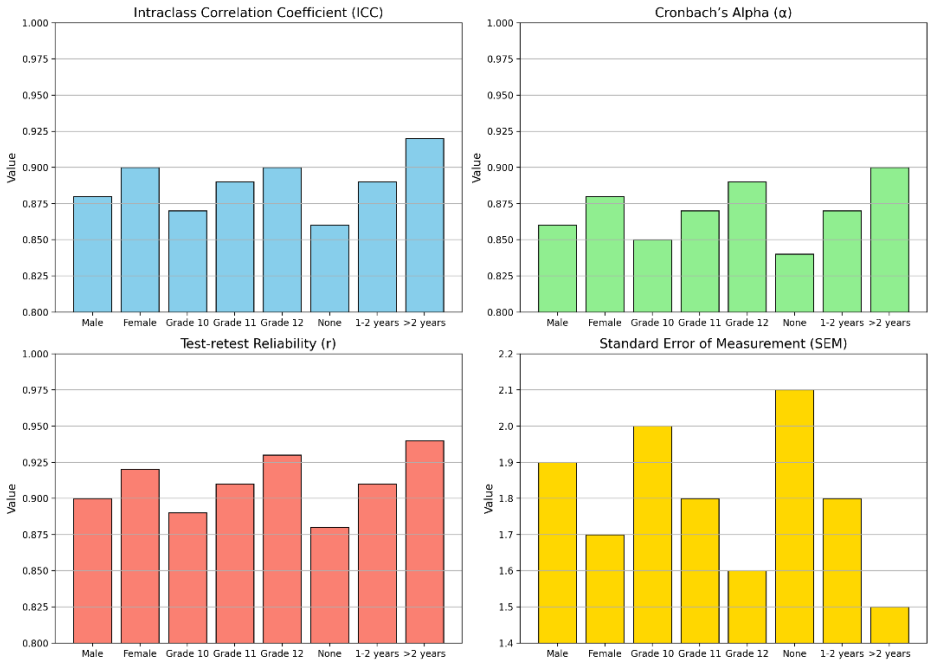

Figure 2. Reliability measure by subgroup and experience level impact on reliability and performance

Table 5 and figure 2 presents the reliability coefficients across different subgroups within our sample of high school students. Analysis of these data reveals several notable patterns in the psychometric properties of our standardized discus throwing assessment instrument.

In terms of gender differences, both male and female subgroups demonstrated excellent reliability, with females showing slightly higher consistency across all reliability metrics (ICC = 0.90, α = 0.88, test-retest r = 0.92) compared to males (ICC = 0.88, α = 0.86, test-retest r = 0.90). The standard error of measurement (SEM) was correspondingly lower for females (1.7) than males (1.9), indicating marginally higher precision in measurements among female participants.

When examining reliability by grade level, a clear progressive pattern emerged. Reliability coefficients systematically increased from Grade 10 (ICC = 0.87, α = 0.85, test-retest r = 0.89) through Grade 11 (ICC = 0.89, α = 0.87, test-retest r = 0.91) to Grade 12 (ICC = 0.90, α = 0.89, test-retest r = 0.93). This trend was mirrored in the decreasing SEM values across grade levels (Grade 10: 2.0, Grade 11: 1.8, Grade 12: 1.6), suggesting that the instrument produces more consistent and precise measurements with older students.

The most pronounced differences in reliability measures were observed across experience levels. Students with no prior discus throwing experience showed good but lower reliability (ICC = 0.86, α = 0.84, test-retest r = 0.88, SEM = 2.1) compared to those with 1-2 years of experience (ICC

= 0.89, α = 0.87, test-retest r = 0.91, SEM = 1.8). Students with more than 2 years of experience demonstrated the highest reliability across all metrics (ICC = 0.92, α = 0.90, test-retest r = 0.94) and the lowest measurement error (SEM = 1.5).

Overall, these subgroup analyses indicate that while our assessment instrument maintains strong reliability across all demographic and experience categories (all ICC values ≥ 0.86), the instrument performs optimally with more experienced athletes, senior students, and female participants. These findings suggest that factors such as movement consistency, technical proficiency, and developmental maturity may influence the reliability of performance assessments in discus throwing among high school athletes.

4. DISCUSSION

The journey to develop and validate a reliable assessment instrument for discus throwing has revealed fascinating insights into both the technical aspects of athletic evaluation and the broader implications for physical education. Our findings paint a compelling picture of how structured assessment tools can bridge the gap between subjective observation and objective measurement in sports performance. The remarkable inter-rater reliability (ICC = 0.89) achieved by our instrument suggests a breakthrough in standardizing what has traditionally been a challenging area of assessment. This achievement becomes particularly meaningful when we consider the complex, dynamic nature of discus throwing – a sport where the subtleties of technique can be as crucial as the final distance achieved (Horst et al., 2020; Chen et al., 2021). As we delved deeper into the component-specific analysis, an interesting pattern emerged. The power position component, achieving the highest reliability (ICC = 0.92), demonstrated how clear biomechanical markers can lead to consistent evaluation even across different raters. This finding challenges the traditional notion that dynamic sport movements are inherently difficult to assess reliably (McKeown et al., 2014). The success in evaluating this critical phase of the throw suggests that when provided with well-defined criteria and proper training, evaluators can achieve remarkable consistency in their assessments. Conversely, the slightly lower reliability in preliminary swings (ICC = 0.86) reveals the ongoing challenge of evaluating more fluid, preparatory movements, pointing to areas where further refinement in assessment criteria might be beneficial.

Perhaps one of the most intriguing aspects of our findings lies in the demographic patterns that emerged from the data. The subtle but consistent difference in reliability between male and female participants (ICC = 0.90 vs 0.88) opens up important questions about gender-specific

considerations in sports assessment. This pattern, while small in magnitude, suggests that our understanding of technical execution and its evaluation might benefit from a more nuanced, gender-aware approach (Gromeier et al., 2017; Sørensen et al., 2022). Similarly, the clear progression in reliability scores across grade levels tells a story of how motor skill development and assessment reliability might be more intertwined than previously thought (Hulteen et al., 2020; Hoeboer et al., 2018).

The experience-level analysis revealed what might be our most significant insight into the relationship between skill development and assessment reliability (Cameron et al., 2016). As participants' experience increased, we observed not only higher reliability scores but also a consistent decrease in measurement error (Yang et al., 2015; Dunbar et al., 1991). This pattern suggests a symbiotic relationship between skill mastery and assessment accuracy – as movements become more refined and consistent, they also become more reliably measurable (Osborne & Gordon, 1972). This insight has profound implications for how we might approach both teaching and assessment in physical education.

Looking toward practical applications, our findings suggest that reliable assessment in physical education need not be a matter of subjective judgment or overly complex measurement systems (Hay, 2006; Starck, 2017). The success of our instrument in achieving high reliability with minimal equipment requirements demonstrates that well-designed assessment tools can be both sophisticated in their approach and practical in their implementation. The ability to complete comprehensive evaluations within a standard class period while maintaining high reliability suggests that quality assessment need not come at the cost of instructional time(Pellegrino & Quellmalz, 2010). Support for sequential skill acquisition models, where more fundamental movement patterns (power position) demonstrate higher reliability than more complex, preparatory movements (preliminary swings) (Rodríguez-Perea et al., 2019; Enokibori & Mase, 2013).

Demonstration of multi-component reliability analysis, addressing calls for more comprehensive approaches to assessment validity (Hulteen et al., 2020). Empirical support for the benefits of standardized assessment instruments in physical education, overcoming challenges of subjective evaluation (Starck, 2017). Insights into demographic considerations in sports performance assessment, expanding theoretical perspectives on fairness and equity (Moura et al., 2020). Support for criterion-referenced assessment in PE, moving beyond norm-referenced approaches. Integration of quantitative and qualitative measures, responding to demands for mixed-method assessment frameworks (Wilkie et al., 2023).

Evidence-based assessment methodology that can inform teacher training and professional development in physical education (Hay, 2006). Opportunity to strengthen alignment between curriculum, instruction, and assessment in PE, a key tenet of effective educational practice (Tsuda et al., 2019). Standardization of performance evaluation, enabling meaningful comparison and benchmarking across settings (Zhu et al., 2011). Support for competency-based education models in physical education, emphasizing mastery over time-based progression (Humphries et al., 2002). These comprehensive findings not only validate the reliability of the developed instrument but also provide valuable insights into the broader context of physical education assessment and skill development evaluation.

5. CONCLUSIONS

This study successfully developed and validated a reliable assessment instrument for discus throwing performance among Indonesian high school students. The instrument demonstrated strong overall inter-rater reliability, with particularly high consistency for the critical Power Position component. These findings indicate that the assessment tool can serve as a practical and efficient means of comprehensively evaluating discus throwing technique, with the potential to significantly enhance skill development, performance tracking, and evidence-based instruction within physical education curricula. The robust reliability of the instrument, especially in assessing the pivotal Power Position, suggests that it can provide educators with an objective, standardised framework to assess and guide students' technical proficiency in this complex athletic skill. This, in turn, could enable more targeted, data-driven approaches to teaching and learning, ultimately supporting the holistic development of discus throwing competency among Indonesian high school students.

6. REFERENCES

1. Alamiri, H. F., & Ameen, M. J. (2023). An Analytical Study of Some Biomechanical Variables and Achieving Fingerprinting in Discus Throwing for Applicants. *European Journal of Sport Sciences, 2*(2), 26-30. https://doi.org/10.24018/ejsport.2023.2.2.71
2. Alhumaid, M. M., & Atta, I. I. (2022). Biomechanical properties of the discus throw: analytical case study of the Paralympic record holder in the F33 category. *Heliyon*, *8*(12), 1-7. https://doi.org/10.1016/j.heliyon.2022.e11926
3. Angell, R. M., Butterfield, S. A., Tu, S., Loovis, E. M., Mason, C. A., & Nightingale, C. (2018). Children's Throwing and Striking: A Longitudinal Study. *Journal of Motor Learning and Development*, *6*(2), 1-30. https://doi.org/10.1123/jmld.2017-0026
4. Cameron, C. E., Cottone, E. A., Murrah, W. M., & Grissmer, D. W. (2016). How Are Motor Skills Linked to Children's School Performance and Academic Achievement? *Child Development Perspectives, 10*(2), 1-14. https://doi.org/10.1111/cdep.12168

5. Chen, C. F., Wu, H. J., Yang, Z. S., Chen, H., & Peng, H. T. (2021). Motion Analysis for Jumping Discus Throwing Correction. *International Journal of Environmental Research and Public Health*, *18*(24), 1-9. https://doi.org/10.3390/ijerph182413414
6. Dai, B., Leigh, S., Li, H., Mercer, V. S., & Yu, B. (2013). The relationships between technique variability and performance in discus throwing. *Journal of Sports Sciences*, *31*(2), 219–228. https://doi.org/10.1080/02640414.2012.729078
7. Dawood, M., Becker, P. J., Rooijen, A. J. V., & Korkie, E. (2018). Inter- and intra-rater reliability of a technique assessing the length of the Latissimus Dorsi muscle. *South African Journal of Physiotherapy*, *74*(1), 1-7. https://doi.org/10.4102/sajp.v74i1.388
8. Dunbar, S.B., Koretz, D., & Hoover, H.D. (1991). Quality Control in the Development and Use of Performance Assessments. *Applied Measurement in Education, 4*(4), 289-303.
9. Enokibori, Y., & Mase, K. (2013). Reproduction of expert's subjective skill-level evaluation with wearable acceleration and gyroscope sensors on metal-filing. *Transactions of the Japanese Society for Artificial Intelligence, 28*(4), 391–399. https://doi.org/10.1527/tjsai.28.391
10. Faber, I. R., Koopmann, T., Büsch, D., & Schorer, J. (2021). Developing a tool to assess technical skills in talented youth table tennis players-a multi-method approach combining professional and scientific literature and coaches' perspectives. *Sports Medicine - Open*, *7*(1), 1-24. https://doi.org/10.1186/s40798-021-00327-5
11. Fajrina, S., Nulhakim, L., & Taufik, A. N. (2021). Pengembangan instrumen performance assessment praktikum untuk mengukur keterampilan proses sains (KPS) siswa SMP kelas VIII pada tema makananku kesehatanku. *PENDIPA Journal of Science Education, 6*(1), 105–112. https://doi.org/10.33369/pendipa.6.1.105-112
12. Gregor, R. J., Whiting, W. C., & McCoy, R. W. (1985). Kinematic analysis of Olympic discus throwers. *International Journal of Sport Biomechanics, 1*(2), 131–138. https://doi.org/10.1123/ijsb.1.2.131
13. Gromeier, M., Koester, D., & Schack, T. (2017). Gender Differences in Motor Skills of the Overarm Throw. *Frontiers in psychology, 8*, 1-12. https://doi.org/10.3389/fpsyg.2017.00212
14. Hay, P. (2006). *Assessment for Learning in Physical Education.* SAGE Publications Ltd eBooks (p. 312). SAGE Publishing. https://doi.org/10.4135/9781848608009.n18
15. Hoeboer, J. J. A. A. M., Ongena, G., Krijger-Hombergen, M., Stolk, E., Savelsbergh, G. J. P., & de Vries, S. I. (2018). The Athletic Skills Track: Age- and gender-related normative values of a motor skills test for 4- to 12-year-old children. *Journal of Science and Medicine in Sport*, *21*(9), 975–979. https://doi.org/10.1016/j.jsams.2018.01.014
16. Horst, F., Janssen, D., Beckmann, H., & Schöllhorn, W. I. (2020). Can Individual Movement Characteristics Across Different Throwing Disciplines Be Identified in High-Performance Decathletes?. *Frontiers in Psychology*, *11*, 1-14. https://doi.org/10.3389/fpsyg.2020.02262
17. Hulteen, R. M., Barnett, L. M., True, L., Lander, N. J., Del Pozo Cruz, B., & Lonsdale, C. (2020). Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. *Journal of Sports Sciences*, *38*(15), 1717–1798. https://doi.org/10.1080/02640414.2020.1756674
18. Humphries, C., Lovdahl, P., & Ashy, M. (2002). Elementary Physical Education and the National Standards. *Journal of Physical Education, Recreation & Dance*, *73*(5), 42–45. https://doi.org/10.1080/07303084.2002.10607808
19. Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, *65*(23), 2276–2284. https://doi.org/10.2146/ajhp070364
20. Kobayashi, Y., Ae, M., Miyazaki, A., & Fujii, N. (2012). Biomechanical characteristics of the overarm throwing motion of skilled elementary school children in comparison with standard

motion models. *Japan Journal of Physical Education, Health and Sport Sciences, 57*(2), 613–629. https://doi.org/10.5432/jjpehss.12032
21. Kobayashi, Y., Ae, M., Miyazaki, A., Fujii, N., Iiboshi, A., & Nakatani, H. (2016). Kinetics of throwing arm joints and the trunk motion during an overarm distance throw by skilled Japanese elementary school boys. *Sports Biomechanics, 15*(3), 314–328. https://doi.org/10.1080/14763141.2016.1161820
22. Liu, H., & Yu, B. (2024). Rotation of the thrower-discus system and performance in discus throwing. *Sports Biomechanics, 23*(5), 624–639. https://doi.org/10.1080/14763141.2021.1880619
23. Marqués-Bruna, P., & Grimshaw, P. (1998). Variability in development of overarm throwing: A longitudinal case study over the first 6 months of throwing. *Perceptual and Motor Skills, 86*(3), 1403–1410. https://doi.org/10.2466/pms.1998.86.3c.1403
24. McKeown, I., Taylor-McKeown, K., Woods, C., & Ball, N. (2014). Athletic ability assessment: a movement assessment protocol for athletes. *International Journal of Sports Physical Therapy, 9*(7), 862–873.
25. Mercier, K., & Doolittle, S. (2013). Assessing student achievement in physical education for teacher evaluation. *Journal of Physical Education, Recreation & Dance, 84*(3), 38–45. https://doi.org/10.1080/07303084.2013.767721
26. Meron, A., & Saint-Phard, D. (2017). Track and Field Throwing Sports: Injuries and Prevention. *Current Sports Medicine Reports, 16*(6), 391–396. https://doi.org/10.1249/JSR.0000000000000416
27. Moura, A., Graça, A., MacPhail, A., & Batista, P. (2020). Aligning the principles of assessment for learning to learning in physical education: A review of literature. *Physical Education and Sport Pedagogy, 26*(4), 388-401. https://doi.org/10.1080/17408989.2020.1834528
28. Osborne, M. M., & Gordon, M. E. (1972). An investigation of the accuracy of ratings of a gross motor skill. *Research Quarterly for Exercise and Sport, 43*(1), 55–62. https://doi.org/10.1080/10671188.1972.10615110
29. Pellegrino, J. W., & Quellmalz, E. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134. https://doi.org/10.1080/15391523.2010.10782565
30. Petr, T. A., & Paskus, T. S. (2009). The collection and use of academic outcomes data by the NCAA. *New Directions for Institutional Research, 2009*(144), 77–92. https://doi.org/10.1002/ir.315
31. Rodriguez-Perea, A., Chirosa Ríos, L. J., Martinez-Garcia, D., Ulloa-Díaz, D., Guede Rojas, F., Jerez-Mayorga, D., & Chirosa Rios, I. J. (2019). Reliability of isometric and isokinetic trunk flexor strength using a functional electromechanical dynamometer. *PeerJ, 7*, 1-17. https://doi.org/10.7717/peerj.7883
32. Sørensen, A., Haugen, E. C., & van den Tillaar, R. (2022). Is There a Sex Difference in Technical Skills among Youth Soccer Players in Norway?. *Sports (Basel, Switzerland), 10*(4), 1-15. https://doi.org/10.3390/sports10040050
33. Spinelli, C. (2007). Introduction: The benefits, uses, and practical application of informal assessment procedures. *Reading & Writing Quarterly, 24*(1), 1–14. https://doi.org/10.1080/10573560701753005
34. Starck, J. R. (2017). Using Assessment for Learning in Sport Education: Column Editor: K. Andrew R. Richards. *Strategies, 30*(4), 51–53. https://doi.org/10.1080/08924562.2017.1321446
35. Thomas, J. R., Alderson, J., Thomas, K. T., Campbell, A., Edwards, W. B., Meardon, S. A., & Elliott, B. (2013). Is there a general motor program for right versus left hand throwing in children? *Journal of Biosensors & Bioelectronics, 4*(3), 1-14. https://doi.org/10.4172/2155-6210.s1-001

36. Trasolini, N. A., Nicholson, K. F., Mylott, J. A., Bullock, G. S., Hulburt, T., & Waterman, B. R. (2022). Biomechanical analysis of the throwing athlete and its impact on return to sport. *Arthroscopy, Sports Medicine, and Rehabilitation, 4*(1), 83–91. https://doi.org/10.1016/j.asmr.2021.09.027
37. Tsuda, E., Wyant, J. D., Bulger, S. M., Elliott, E., Taliaferro, A. R., Burgeson, C. R., & Wechsler, H. (2019). Recommendations for developing and implementing state-level physical education accountability systems in student learning. *Journal of Physical Education, Recreation & Dance, 90*(9), 9–15. https://doi.org/10.1080/07303084.2019.1658497
38. Wilkie, B., Jordan, A., Foulkes, J., Woods, C. T., Davids, K., & Rudd, J. (2023). Examining the validity, reliability and feasibility of capturing children's physical literacy through games-based assessment in physical education. *Frontiers in Sports and Active Living, 5,* 1-15. https://doi.org/10.3389/fspor.2023.1188364
39. Yang, Y., Oosterhof, A., & Xia, Y. (2015). Reliability of scores on the summative performance assessments. *The Journal of Educational Research, 108*(6), 465–473. https://doi.org/10.1080/00220671.2014.917255
40. Yarmani, Y., Nopiyanto, Y. E., & Noprizal, H. (2021). Effort to increase interest and learning outcome of discus throwing by using melamine plate media. *Kinestetik: Jurnal Ilmiah Pendidikan Jasmani, 5*(2), 411–418. https://doi.org/10.33369/jk.v5i2.17787
41. Young, R. W. (2009). The ontogeny of throwing and striking. *Human Ontogenetics, 3*(1), 19–33. https://doi.org/10.1002/huon.200800013
42. Yu, B., Broker, J., & Silvester, L. J. (2002). Athletics: A kinetic analysis of discus-throwing techniques. *Sports Biomechanics, 1*(1), 25–45. https://doi.org/10.1080/14763140208522785
43. Zhu, W., Rink, J. E., Placek, J. H., Graber, K. C., Fox, C., Fisette, J. L., Dyson, B., Park, Y., Avery, M., Franck, M., & Raynes, D. (2011). PE metrics: Background, testing theory, and methods. *Measurement in Physical Education and Exercise Science, 15*(2), 87–102. https://doi.org/10.1080/1091367x.2011.568363

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

COPYRIGHT