Journal Sport Area
http://journal.uir.ac.id/index.php/JSP
Vol. 10. No. 3. December, (2025)

Development and validation of an instrument to assess attacking tactical behavior in football

Firmansyah Dahlan[1abcde,*], Fajri Hamzah[1cde], Binsar Hasudungan Sihombing[1cde], Ahmad Syarif Burhan[1cde], & Junaidi[2acde]

Universitas Muhammadiyah Palopo, Indonesia[1]
Rajamangala University of Technoloby Krunghtep, Thailand[2]

Received 15 July 2025; Accepted 13 December 2025; Published 19 December 2025
Ed 2025; *10*(3): 490-503

ABSTRACT

Background: Existing tactical assessment tools generally evaluate overarching game principles but lack positional specificity and do not integrate decision-making and execution within a coherent framework. Moreover, these instruments were developed outside the Indonesian context, creating a clear gap for a position-integrated tool tailored to the tactical demands of youth football in Indonesia. **Objectives**: Thusstudy aimed to develop and validate a new instrument for assessing attacking tactical behavior in youth football by incorporating positional roles and experience-level differences. **Methods**: The instrument was developed through content validation with seven experts using Aiken's V. Construct validity was subsequently assessed via the known-groups method, involving 72 football players from a university club and youth teams. The instrument's reliability was further evaluated through inter-rater reliability with three raters and test–retest reliability with 36 youth football players. **Finding/Results**: The instrument demonstrated excellent content validity (Aiken's V = 0.952) and good discriminant validity, with college players scoring significantly higher than youth players. Reliability analyses indicated strong inter-rater (ICC = 0.837-0.939) and test–retest reliability (ICC = 0.801-0.923). Compared with other instruments, the tool offers greater positional detail and holistic integration of tactical behavior, enabling more context-specific evaluation of attacking actions. **Conclusion**: The instrument was found to be valid and reliable for assessing attacking tactics in youth football. Although the sample size was limited and the instrument development did not yet address defensive and transitional phases, these findings support further refinement and application in future research.

Keywords: Tactical behaviors; tactical action; offensive football; test instruments

https://doi.org/10.25299/sportarea.2025.vol10(3).23989

Corresponding Author: Firmansyah Dahlan, Department of Physical Education, Faculty of Teacher Training and Education, Universitas Muhammadiyah Palopo, Palopo, Indonesia
firmansyahdahlan@umpalopo.ac.id

How to Cite: Dahlan, F., Hamzah, F., Sihombing, B. H., Burhan, A. S., & Junaidi. (2025). Development and validation of an instrument to assess attacking tactical behavior in football. *Journal Sport Area*, *10*(3), 490-503. https://doi.org/10.25299/sportarea.2025.vol10(3).23989

Authors' Contribution: a – Study Design; b – Data Collection; c – Statistical Analysis; d – Manuscript Preparation; e – Funds Collection

INTRODUCTION

Physical, technical, mental, tactical, and strategic components are widely recognized as key determinants of sports performance (Bompa & Buzzichelli, 2019; Balyi et al., 2013), including football (FIFA, 2021;

Dahlan et al., 2020). However, a recent review of sports journals indexed in SINTA shows that research on football tactics and strategy remains very limitedonly 2% of published articlescompared with technique (36%), physical conditioning (31%), and mental aspects (7%) (Dahlan et al., 2024). This indicates that studies on football tactics, especially research involving players and grounded in Indonesian football principles, are still insufficient. Strategy refers to the art of planning, while tactics represent the mechanisms used to execute those plans (Bompa & Buzzichelli, 2019; Balyi et al., 2013). In football, tactics operate across three game cycles: defense, attack, and transition (defense to attack, and attack to defense) (Dahlan et al., 2024).

Tactical behaviour involves determining when, where, and why physical, technical, and mental skills are used (Snow & Thomas, 2009), and reflects the continuous exchange of information between players (Folgado et al., 2018). Tactical behaviors are abilities that are really needed when there is interaction between players (one team and the opponent) (Olthof et al., 2015). It includes the ability to interpret rules and demands of the game phase (Bach Padilha et al., 2017), and to read affordances (opportunities) emerging from game situations (Silva et al., 2013). Tactical actions, such as dribbling, passing, and shooting, emerge from affordances related to space and time created by teammates and opponents (Laakso et al., 2022). They also involve managing space through players' positions and movements (Correia da Silva et al., 2021). Team behavior, individual actions (offensive and defensive), and ball possession effectiveness can be predictors of tactical actions assessment (Laakso et al., 2022). Football tactics encompass two distinct but integrated concepts: tactical behavior and tactical action. Tactical behavior refers to a player's ability to read game situations and process information to identify potential opportunities. In contrast, tactical action represents the practical execution of tactical behavior, including technical skills such as dribbling, passing, and shooting, as well as spatial management through positioning and movement during the game.

Tactical training is the most researched aspect, even the treatment tends to be the same, namely small-sided conditioned games (Dahlan et al., 2024). As for tests, assessments, and/or tactical evaluations, they have been very minimally researched. Football tactical test instruments in the 1980s-1990s were still largely ignored, even the test instruments that are considered capable of representing football tactics to date are only six instruments; Game performance assessment instrument (GPAI), Performance assessment in team sports (TSAP), Procedural tactical knowledge test (KORA), Game performance evaluation tool (GPET), System of tactical assessment in soccer (FUT-SAT), and Game performance analysis (González-Víllora et al., 2015). While international tools such as FUT-SAT, TSAP, and GPET have been widely applied in Europe and South America, research on tactical-performance assessment in Asian football, particularly in Indonesia, remains limited. Existing domestic studies primarily focus on technical skills or physical conditioning, with no validated instruments specifically designed to evaluate attacking tactical behavior in Indonesian youth football. This gap is critical given differences in training culture, match exposure, and tactical education compared with countries where most tactical instruments were developed.

Based on these limitations, this study aims to develop a football tactical test instrument tailored to the attacking phase. In this study, tactical behaviour refers to players' decision-making and movement patterns used to create or exploit advantages, while tactical action represents the observable execution of these decisions. Both components are integrated within a single assessment framework. The development process follows standard instrument design procedures, including defining variables, determining indicators, and establishing operational procedures (Muñiz & Fonseca-Pedrero, 2019; Ferrando et al., 2022). The operational test procedures adapt elements from GPAI, TSAP, KORA, GPET, and FUT-SAT, including small-sided game formats (4 + 1 vs. 4 + 1 on a 50 × 30 m field; 3 + 1 vs. 3 + 1 on a 36 × 27 m field), duration protocols (2 × 4 minutes with a 3-minute interval), and standard football rules except the offside rule. These procedures closely mirror real match conditions, reducing the risk of overlooking talented players due to non-ecological testing environments (Nunes et al., 2020). The next adopted procedure is the assessment of GPAI, TSAP, KORA, GPET, and FUT–SAT, namely submitting indicators, giving assessments if the indicators are achieved, assisted by the assessment team to observe the achievement of the indicators, the observation process is carried out during the game and/or analysis of video recordings of the test implementation. The context that we elaborate on is the indicators used by GPAI, TSAP, KORA, GPET, and FUT-SAT and the latest research recommendations on football attack predictors.

This study introduces two key innovations. First, player compositions must match positional roles (defenders, midfielders, strikers) to reflect position-specific tactical demands, as players in the same position exhibit identifiable tactical patterns (Guard et al., 2022). For example, when implementing a 4 + 1 Goalkeeper (GK) vs. 4 + 1 (GK) format, the player composition should be derived from identical positional roles 2 defenders (DF), 1 midfielder (MF), and 1 striker (ST) vs. 2 defenders (DF), 1 midfielder (MF), and 1 striker (ST). Second, the instrument accounts for differences in competitive experience, as experience influences tactical performance, motivation, and competitive behaviour (Slimani et al., 2016). Examples include matching regular players against regular players, non-regular players against non-regular players, or mixed groups of regular and non-regular players. Another innovation is the integration of digital platforms to support the observation process. The study also integrates Tracker Video Analysis® (Open Source Physics (OSP), National Science Foundation and Davidson College, US), a digital platform suitable for analyzing movement and reducing observational errors (Buldú et al., 2019; Castillo et al., 2020; Custódio et al., 2022; Narizuka et al., 2023), and widely used in physical education contexts (Artiningsih & Nurohman, 2020; Muliyati et al., 2020). The target users of the instrument are players aged 14-17 years, an age group where regular training and tactical understanding become essential (FIFA, 2021).

In summary, unlike established tools that assess tactics through general game situations, the instrument developed in this study integrates positional specificity, attacking-phase context, experience-based player grouping, and video-assisted evaluation. These components address methodological gaps in existing literature and support the development of a valid and reliable assessment tool for attacking tactical behaviour in youth football.

METHOD

Research Design

This study employed a research and development (R&D) approach to construct and validate an observational instrument for assessing attacking tactical behavior in football. The development process followed a modified 3D model, comprising the Define, Design, and Develop stages, which are commonly used in instrument construction and validation studies. The original 4D development model proposed by Thiagarajan et al. (1974) was adapted into a 3D model by omitting the Disseminate stage. This modification was justified because the instrument currently assesses only the tactical performance of youth players, excluding all other stages of athlete development and adult-elite players, as well as the defensive and transition phases. Moreover, the participants involved (both experts and players) were exclusively from Indonesia. Therefore, the instrument has not yet reached the dissemination stage and is not yet prepared for large-scale implementation.

Participants

Content validation was conducted by a panel of seven experts, including academics and football practitioners. Following expert validation, construct validity was assessed using the known-groups method with 36 university-level and 36 youth football players. Instrument reliability was then examined through inter-rater reliability with three evaluators and test–retest reliability with 36 youth players. Youth participants aged 14-15 years were recruited via purposive sampling from football academies in South Sulawesi Province; the same approach was applied to university-level players to ensure active engagement in structured training and competitive matches.

For construct validation and test–retest reliability, players were evenly divided into twelve balanced teams based on positional roles and playing experience. Games followed a 3 + 1 (goalkeeper) versus 3 + 1 format on a 36 × 27 m field, lasting 2 × 4 minutes (see **Table 1**). Six evaluators each observed one player per match. Before testing, both players and evaluators received briefings on the procedures (Table 1) and the domain–indicator blueprint (Table 2) to ensure smooth implementation. For inter-rater reliability, three evaluators assessed performance using Tracker Video Analysis®. All participants provided informed consent. The study received ethical approval from the Ethics Committee of Universitas Muhammadiyah Palopo, Indonesia (No. 01/KEP/III.3.AU/F/2024).

Instrument Development

The instrument was designed to assess attacking tactical behavior through observable actions in small-sided games. Development involved identifying relevant domains (decision-making, execution, positional relevance), generating indicators, designing operational procedures, and establishing scoring criteria. All matches were recorded and analyzed using a standardized video-analysis application to ensure accuracy and consistency across evaluators.

Content and Construct Validity

Content validity was evaluated by seven experts (football coaches, sport scientists, and tactical practitioners) who rated each indicator's relevance and clarity on a 5-point scale. Aiken's V coefficient determined expert agreement, with $V \geq 0.70$ as the minimum acceptable threshold. Indicators below this cut-off were revised. The overall Aiken's V was 0.952, indicating excellent content validity and strong consensus. Construct validity was assessed using the known-groups method to verify the instrument's ability to differentiate between groups with differing tactical expertise (e.g., novice vs. advanced players). Significant score differences confirmed the instrument's discriminative capability, supporting its construct validity. These stages collectively provide robust evidence that the instrument accurately measures attacking tactical behavior and differentiates between skill levels.

Reliability Testing

Reliability was examined in two stages. First, inter-rater reliability assessed agreement among independent evaluators. Given the subjective nature of tactical assessment, this step ensures consistency across raters. Evaluators scored independently, and agreement was quantified using the inter-rater ICC (e.g., ICC 2,k) or equivalent. ICC values were interpreted as poor (< 0.50), moderate (0.50-0.75), good (0.75-0.90), or excellent (> 0.90). High inter-rater ICC indicated stable, reproducible scores. Second, test-retest reliability was evaluated using ICC 3,1 to assess stability over time, considering both consistency and absolute agreement. Using the same thresholds, results indicated the instrument produces consistent scores across repeated measurements. Together, inter-rater and test-retest analyses confirm that the instrument is reliable, providing stable and reproducible results suitable for research and applied practice.

Data Analysis

All statistical analyses were performed using standard software (e.g., SPSS or R). Content validity was assessed using Aiken's V, with a threshold of ≥ 0.70 indicating acceptable agreement among experts. Construct validity was examined using the known-groups method by comparing total scores between youth and university-level players via independent t-tests. Inter-rater and test-retest reliability were evaluated using Intraclass Correlation Coefficients (ICC 2,k and ICC 3,1, respectively), interpreted as: < 0.50 = poor, 0.50-0.75 = moderate, 0.75-0.90 = good, > 0.90 = excellent. Descriptive statistics (mean ± SD, minimum, maximum) were reported for total scores. Statistical significance was set at $p < 0.05$.

RESULTS AND DISCUSSION

Define

The analysis of needs related to football tactics and strategies remains highly limited and is rarely examined in depth. Existing tactical assessment tools such as GPAI, TSAP, GPET, KORA, and FUT-SAT evaluate general game principles but lack positional specificity and do not integrate decision-making and execution into a single, coherent framework. Moreover, these instruments were developed outside the Indonesian context, leaving a clear gap for a position-integrated tool tailored to the tactical demands of youth football in Indonesia. The test instrument developed in this study offers several advantages, including a more specific and structured measurement of tactical behavior, the integration of technical actions and spatial management, and the consideration of key factors that differentiate tactical performance such as match experience and positional differences. In addition, the use of digital technology in evaluating tactical performance can reduce assessment errors that may arise from direct visual observation.

Design

Product prototype;

Table 1. Implementation Procedure of the Football Attacking Tactical Behaviors Test Instrument

Section	Description
Purpose of the Test	To measure tactical behavior demonstrated through tactical actions such as dribbling, passing, shooting, the ability to manage space through position changes, and player interaction during the attacking phase.
Target Users	Soccer players aged 14-17 years.
Game Format	• 4 + 1 (GK) vs 4 + 1 GK (Goalkeeper) on a 50 x 30 m field. • 3 + 1 (GK) vs 3 + 1 (GK) on a 36 x 27 m field
Game Duration	2 x 4 minutes or 2 x 3 minutes intervals.
Rules	All standard football rules apply except off-side.
Player Distribution	Players must be evenly distributed: GK (Goalkeeper), DF (Defender), MF (Midfielder), ST (Striker) vs GK, DF, MF, ST.Examples:Regular DF vs Regular DF on the opposing team. If one team has 4-3 DF, the opposing team must also have 4-3 DF.
Assessment Procedures	• Assessment team: 6-8 individuals, with one assessor assigned per player. • Assessors must be football experts and/or experienced. • A camera should record the game from a position covering the entire field. • The video can be analyzed using the Tracker Video Analysis® application. • Prior to testing, players and assessors were briefed on the procedures and conducted a trial run of the implementation.

Table 2. Assessment Process of the Football Attacking Tactical Behaviors Test Instrument

Domain	Indicator	Sub-Indicator	Scoring
Action	Option Selection	• Passing is effective to start and/or accelerate an attack. • Key passes to produce assists and or on-target shooting opportunities. • Dribbling is effective in breaking down tight pressing of opponents to get the ball moving. • Key dribbling leads to opportunities and/or fouls in the opponent's defensive area. • Effective shooting results in a goal and/or on target.	Each indicator achieved by a player is assigned a value of "1". Repeated actions of the same indicator are accumulated. The total of all achieved indicators becomes the player's final score.
Positioning	Space Awareness	• Moves into open play space and successfully receives the pass. • Move to open up space to continue the opponent's pressing (even if not receiving a pass).	
	Offensive Coverage	• Winning the ball and creating a defensive to offensive transition.	
	Penetration	• Making an attack attempt and being blocked by opposing players (free kicks, corner kicks, and/or throw-ins). • Initiating an attack that leads to a foul on the opposing player (free kick, corner kick, and/or throw-in).	

Develop

The prototype submitted in the previous stage was subsequently validated by experts. The expert validation results consisted of assessment scores and improvement suggestions. The most notable suggestion was the addition of a new provision: the age of the participants involved in the testing should be classified into two groups, 14-15 years and 16-17 years, considering that these age ranges exhibit significantly different levels of ability. As presented in **Table 3**, the Aiken index calculation results indicate that, out of seven experts, five experts rated the prototype as 4 (very relevant) and two experts rated it as 3 (relevant), resulting in an Aiken validity coefficient of 0.952, which can be interpreted as "valid".

Table 3. Expert Validation Results

Question	Evaluator (Expert)							Aiken Validity	Information
	1	2	3	4	5	6	7		
How is the correspondence between the tested variables and the developed instruments?	4	4	4	3	3	4	4	0.952	Valid

Table 4 presents the results of the descriptive statistics (mean ± SD, minimum, and maximum) obtained during the assessment of construct validity using the known-groups method. Inter-rater and test–retest reliability were evaluated using Intraclass Correlation Coefficients (ICC 2,k and ICC 3,1, respectively).

Table 4. Descriptive Statistics

Test / Group / Time	N	Mean ± SD	Minimum	Maximum
Known-groups: Youth (14–15 yrs)	36	18.72 ± 4.937	10	29
Known-groups: University	36	30.03 ± 4.632	20	40
Test-retest: Time 1 (Youth)	36	17.72 ± 5.057	8	25
Test-retest: Time 2 (Youth)	36	18.22 ± 5.977	4	29
Test-retest: Time 3 (Youth)	36	20.11 ± 4.695	13	29
Intra-rater: Evaluator 1	36	17.78 ± 4.969	8	25
Intra-rater: Evaluator 2	36	18.58 ± 5.250	9	29
Intra-rater: Evaluator 3	36	20 ± 4.864	11	29

Based on the **Table 5**, The validity of the instrument was evaluated using the known-groups method by comparing scores between college football team players and youth football players. College football team players scored significantly higher than youth football players ($t = 10.02$, $df = 70$, $p < 0.001$; mean difference = 11.31, 95% CI: 9.06–13.56; Cohen's $d = 2.35$), and Levene's test confirmed homogeneity of variances ($F = 0.043$, $p = 0.44$). These findings support the discriminant validity of the instrument in distinguishing groups with different characteristics.

Table 5. Known-Groups Validity

Variable	Levene's F	Levene's p	T	df	p (2-tailed)	Mean Difference	95% CI Lower	95% CI Upper
Result	0.04306	00.44	10.02	70	< 0.001	11.31	09.06	13.56

Based on the **Table 6**, the inter-rater reliability of the instrument was assessed using a two-way random-effects model with absolute agreement. The ICC (2,1) for individual scores (single measures) was 0.837 (95% CI: 0.682-0.917, $p < 0.001$), indicating good reliability. For average scores (average measures), the ICC (2,k) was 0.939 (95% CI: 0.865-0.971, $p < 0.001$), indicating excellent reliability when the final score is calculated as the mean of multiple raters. These results demonstrate that the instrument exhibits high consistency across different raters.

Table 6. Inter-Rater Reliability

Source	ICC	95% CI Lower	95% CI Upper	F	df1	df2	Sig. (p)
Single Measures	0.801[a]	0.633	0.895	17.084	35	70	0
Average Measures	0.923[c]	0.838	0.962	17.084	35	70	0

Based on the **Table 7**, the test–retest reliability of the instrument was assessed using a two-way mixed-effects model with absolute agreement. The ICC (3,1) for individual scores (single measures) was 0.801 (95% CI: 0.633-0.895, $p < 0.001$), indicating good reliability. For average scores (average measures), the ICC (3,k) was 0.923 (95% CI: 0.838-0.962), indicating excellent reliability when the final score is calculated as the mean of multiple measurements. These results indicate that the instrument shows good consistency over three test–retest occasions.

Table 7. Test-Retest Reliability

Source	ICC	95% CI Lower	95% CI Upper	F	df1	df2	Sig. (p)
Single Measures	0.801[a]	0.633	0.895	17.084	35	70	0
Average Measures	0.923[c]	0.838	0.962	17.084	35	70	0

After establishing the instrument's validity and reliability, we conducted assessments with 36 youth players to examine the effects of positional differences and variations in player capacity or experience (e.g., regular versus non-regular players). **Figure 1** presents the distribution of tactical performance across player positions Defender (DF), Midfielder (MF), and Striker (ST). **Figure 2** compares the tactical performance of regular and non-regular DF players, while **Figure 3** provides the corresponding comparison for MF players. **Figure 4** illustrates the differences in tactical performance between regular and non-regular ST players.

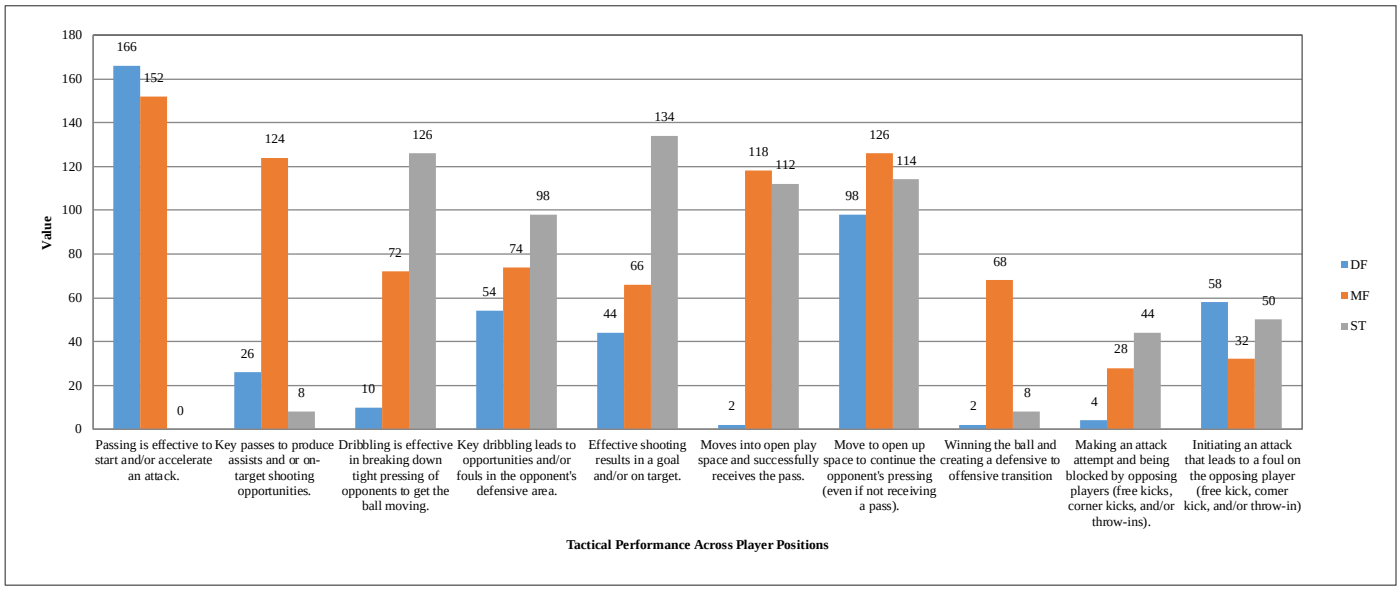

Figure 1. Attacking Tactical Performance Patterns Based on Player Positions

Figure 1 shows that DF is dominant in indicator 1 and indicator 10. MF is dominant in indicators 2, 6, 7, and 8. ST is dominant in indicators 3, 4, 5, and 9. So it is very appropriate if the player's position factor is distributed evenly during the test, because each position has different dominant tactical performance and even has very extreme differences.

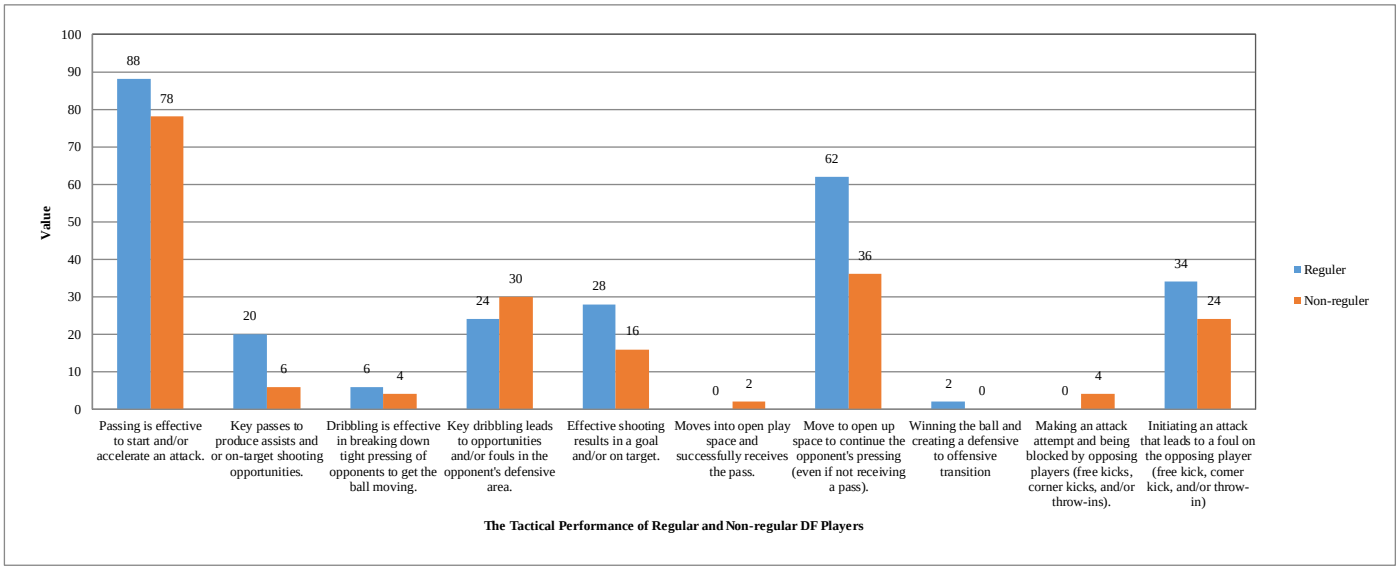

Figure 2. Tactical Performance of Regular and Non-regular DF Attacking

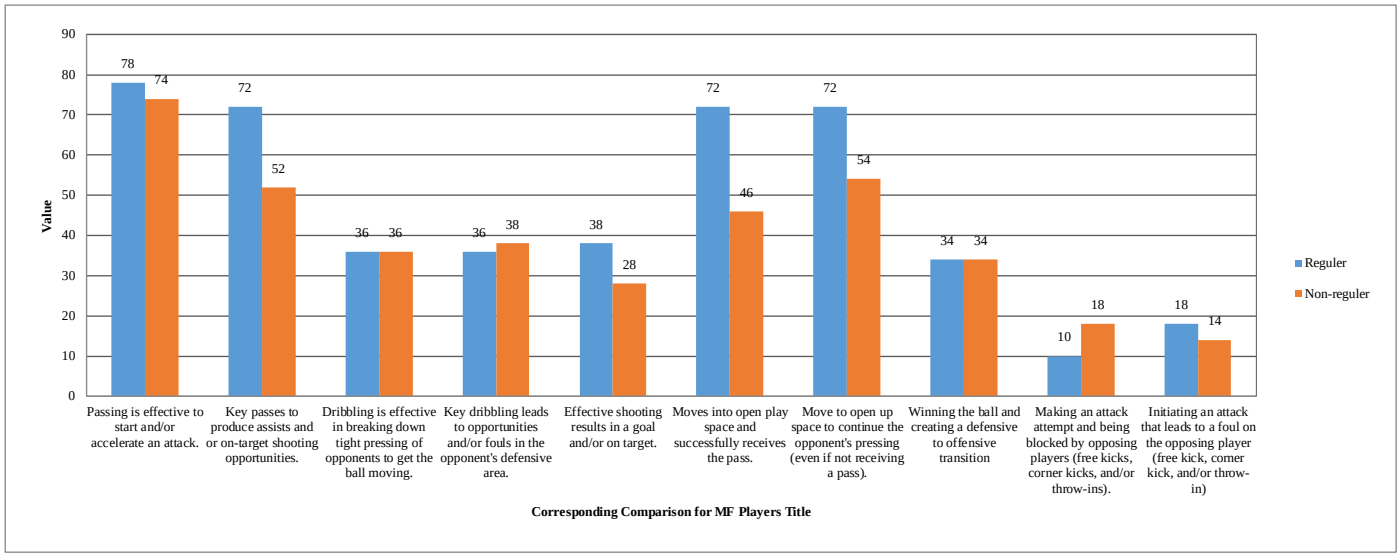

Figure 3. Tactical Performance of Regular and Non-regular MF Attacks

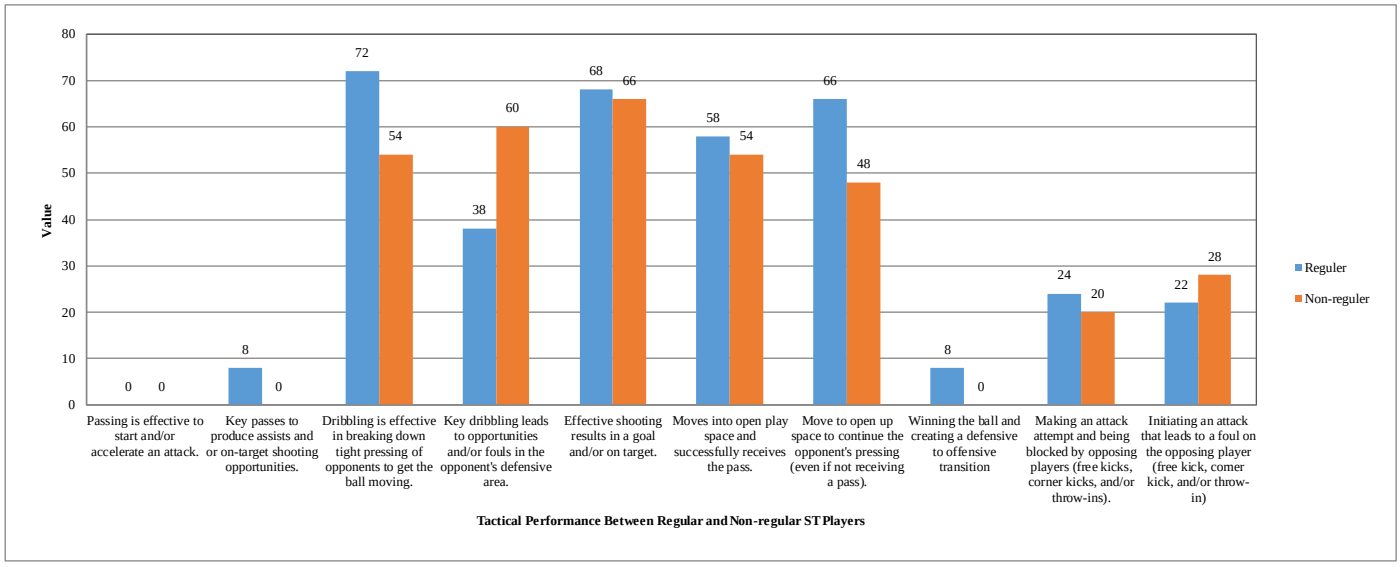

Figure 4. Tactical Performance of Regular and Non-regular ST Attacks

Figure 2, **Figure 3**, **Figure 4**, show that the tactical performance of attacking football DF, MF, and ST from regular players is better than non-regular players. So it is very appropriate if the playing minutes and or playing experience factors of players (regular and non-regular) are distributed evenly during the test, because it has an effect on the difference in tactical performance of attacking a football player.

The findings of this study indicate that the developed performance test instrument possesses strong content validity, as reflected by an Aiken's V coefficient of 0.952, categorized as "very valid" based on expert evaluations. This high coefficient suggests that all test items are relevant and representative of the targeted construct (Yudiana et al., 2017; Milovanovic, 2023). This outcome likely stems from the theoretical and methodological assumptions applied during instrument development, including the integration of decision-making and motor execution within a coherent framework, position-specificity, and grouping based on player experience (regular vs. non-regular players). Strong content validity provides an essential foundation for subsequent construct validation procedures, supporting ongoing empirical refinement and strengthening of the instrument (Sequeira et al., 2024; Umar & Lian, 2024). Construct validity assessed using the known-groups method demonstrated good discriminant validity, with collegiate-level players scoring significantly higher than younger players. These findings align with previous studies using similar methods, which indicated that teams of differing performance levels should yield different results (Conte et al., 2019; Boddington et al.,

2019). Moreover, these results are consistent with prior evidence showing that champion-labeled players and teams exhibit more varied tactical performance, distinguishing them from lower-level players and teams (Correia da Silva et al., 2019). Following the validation of content and construct, the next step was to assess reliability. Skill assessment instruments must demonstrate acceptable validity and adequate reliability before practical application (Boddington et al., 2019). Reliability analyses revealed strong inter-rater reliability (ICC = 0.837-0.939) and high test-retest reliability (ICC = 0.801-0.923). The high inter-rater reliability indicates that the instrument is objective, epistemically defensible, and ensures consistent interpretation of the construct, which is critical given the complex and dynamic nature of decision-making and tactical execution in football (Práxedes et al., 2019; Brito E Sousa et al., 2019; Caso & van der Kamp, 2020; Teoldo da Costa & de Freitas Silvino, 2023). The observed reliability falls within the good to excellent category and aligns with psychometric standards, demonstrating the instrument's ability to produce stable assessments across different raters (Hancock et al., 2025; Thai et al., 2025).

These findings are consistent with existing sports literature emphasizing the importance of inter-rater reliability in athlete performance evaluation. For example, the Landing Error Scoring System (LESS) demonstrated excellent ICC values for both experienced and novice raters assessing professional football players' landing mechanics, highlighting that clear, standardized instruments can yield high inter-rater agreement (Liveris et al., 2024). Other studies on the reliability of fidelity assessments in youth football training indicate that instruments with standardized criteria and explicit guidelines can achieve substantial inter-rater agreement (Kappa ≈ 0.80), even in contexts not directly assessing tactical behaviors (Ljunggren et al., 2019). Nonetheless, inter-rater reliability is influenced by rater experience and rubric clarity, as more experienced raters tend to show higher consistency (Liveris et al., 2024). Therefore, comprehensive rater training and detailed scoring guidelines are recommended to maintain high reliability across various assessment settings. In addition to inter-rater reliability, test-retest reliability evaluates the instrument's stability over time. Test-retest assesses score consistency when the same participants are tested at two different time points, such that score variation primarily reflects measurement error rather than actual behavioral change (Eriksrud & Westheim, 2025). In this study, test-retest ICC values ranging from 0.801 to 0.923 among 36 players indicate strong temporal stability, consistent with sports reliability standards, where ICC ≥ 0.75 is considered good to excellent (Grgic et al., 2021). Other football studies similarly report high test-retest reliability for skill and fitness assessments, such as the FIFA-F-MARC battery, reinforcing that clear and consistent measurement procedures yield stable scores over time (Padrón-Cabo et al., 2019). High test-retest reliability ensures that the instrument is suitable for longitudinal research and intervention evaluation, as score differences more accurately reflect actual changes in tactical behavior rather than random fluctuations or testing conditions (Osborne et al., 2025; Harper et al., 2016). Consequently, the instrument demonstrates both construct validity and reliability, making it a robust tool for evaluating tactical performance in youth football.

Analysis related to Figure 1 shows that player position serves as a differentiating factor in tactical performance outcomes. Players' roles (defender, midfielder, attacker) require distinct technical, tactical, and physical capabilities, which should be considered in training and match play (Laakso et al., 2022). Moreover, positional effects extend to teammates, as tactical behaviors emerge through interactions among team members and opponents, mediated by information exchange based on individual physical, technical, and tactical abilities (Olthof et al., 2015; Folgado et al., 2018). **Figures 2-4** demonstrate that team composition (player position and regularity) influences individual and team tactical performance (Guard et al., 2022; Slimani et al., 2016), while self-efficacy, influenced by experience, also affects athlete performance (Kwon et al., 2022). This study contributes to understanding tactical performance in youth football by introducing an instrument that integrates decision-making and motor execution within a position-specific framework, capturing attacking tactical behaviors more contextually than existing tools. Differences across player positions and experience levels highlight how tactical expertise emerges from interactions between individual abilities and game demands. The instrument demonstrated excellent content validity, discriminant validity, and strong inter-rater and test–retest reliability, supporting its use as both an evaluative and conceptual tool. Limitations include a small, homogeneous sample and exclusive focus on attacking actions, with potential variability due to rater experience and video-based assessments. Future research should expand the sample,

include defensive and transitional phases, implement standardized rater training, and explore automated video analysis to enhance objectivity and practical applicability.

CONCLUSION

This study successfully developed and validated a novel instrument for assessing attacking tactical behaviour in youth football, integrating decision-making and motor execution within a position-specific framework. The instrument demonstrated excellent content validity (Aiken's V = 0.952), discriminant validity, and strong inter-rater and test–retest reliability (ICC = 0.801-0.939), confirming its robustness as both a conceptual model and a practical evaluative tool. The findings emphasize the importance of positional roles and player experience in shaping tactical performance, providing empirical support for ecological dynamics and highlighting the need to consider individual and situational interactions in tactical assessment. Practically, the instrument can guide coaches and practitioners in designing evidence-based training programs, evaluating performance more accurately, and tracking developmental progress in youth football. Despite its strengths, the study is limited by a relatively small and homogeneous sample and its focus solely on attacking actions, excluding defensive and transitional phases. Inter-rater reliability may also vary depending on evaluator experience, and video-based assessment could be influenced by contextual factors. Future research should expand the sample, incorporate all game phases, implement standardized rater training, and explore automated video analysis to further enhance objectivity and applicability. Overall, this instrument provides a valid, reliable, and practical foundation for advancing tactical assessment and evidence-based coaching in youth football.

ACKNOWLEDGEMENTS

The author would like to thank the support of the Universitas Muhammadiyah Palopo which has helped run this research smoothly.

CONFLICT OF INTEREST

All authors have no conflict of interest regarding this article.

REFERENCES

Artiningsih, A., & Nurohman, S. (2020). Analysis of Investigative Skills Based on the Use of Tracker Video Analysis for 21th Century Skill. *Journal of Science Education Research*, *3*(2), 81–86. https://doi.org/10.21831/jser.v3i2.30621

Bach Padilha, M., Guilherme, J., Serra-Olivares, J., Roca, A., & Teoldo, I. (2017). The Influence of Floaters on Players' Tactical Behaviour in Small-Sided and Conditioned Soccer Games. *International Journal of Performance Analysis in Sport*, *17*(5), 721–736. https://doi.org/10.1080/24748668.2017.1390723

Balyi, I., Way, R., & Higgs, C. (2013). *Long-Term Athlete Development*. Human Kinetics.

Boddington, B. J., Cripps, A. J., Scanlan, A. T., & Spiteri, T. (2019). The Validity and Reliability of The Basketball Jump Shooting Accuracy Test. *Journal of Sports Sciences*, *37*(14), 1648–1654. https://doi.org/10.1080/02640414.2019.1582138

Bompa, T. O., & Buzzichelli, C. A. (2019). *Periodization: Theory and Methodology of Training* (6th ed.). Human Kinetics.

Brito E Sousa, R., Bredt, S. D. G. T., Greco, P. J., Manuel Clemente, F., Teoldo, I., & Praça, G. M. (2019). Influence of Limiting The Number of Ball Touches on Players' Tactical Behaviour and Network Properties during Football Small-Sided Games. *International Journal of Performance Analysis in Sport*, *19*(6), 999–1010. https://doi.org/10.1080/24748668.2019.1689751

Buldú, J. M., Busquets, J., Echegoyen, I., & Seirul.lo, F. (2019). Defining a Historic Football Team: Using Network Science to Analyze Guardiola's F.C. Barcelona. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-49969-2

Caso, S., & van der Kamp, J. (2020). Variability and Creativity in Small-Sided Conditioned Games among Elite Soccer Players. *Psychology of Sport and Exercise*, *48*(May), 101645. https://doi.org/10.1016/j.psychsport.2019.101645

Castillo, D., Raya-González, J., Manuel Clemente, F., & Yanci, J. (2020). The Influence of Youth Soccer Players' Sprint Performance on the Different Sided Games' External Load Using GPS Devices. *Research in Sports Medicine*, *28*(2), 194–205. https://doi.org/10.1080/15438627.2019.1643726

Conte, D., Smith, M. R., Santolamazza, F., Favero, T. G., Tessitore, A., & Coutts, A. (2019). Reliability, Usefulness and Construct Validity of The Combined Basketball Skill Test (CBST). *Journal of Sports Sciences*, *37*(11), 1205–1211. https://doi.org/10.1080/02640414.2018.1551046

Correia da Silva, D., Teoldo da Costa, V., Casanova, F., Manuel Clemente, F., & Teoldo, I. (2019). Comparison between Teams of Different Ranks in Small-Sided and Conditioned Games Tournaments. International Journal of Performance Analysis in Sport, 19(4), 608–623. https://doi.org/10.1080/24748668.2019.1643598

Correia da Silva, D., Lopes, M. C., González-Víllora, S., Sarmento, H., & Teoldo, I. (2021). Tactical Behaviour Differences of High and Low-Performing Youth Soccer Players in Small-Sided and Conditioned Games. *International Journal of Performance Analysis in Sport*, *21*(1), 33–50. https://doi.org/10.1080/24748668.2020.1843214

Custódio, I. J. de O., Praça, G. M., Paula, L. V. de, Bredt, S. da G. T., Nakamura, F. Y., & Chagas, M. H. (2022). Intersession Reliability of GPS-Based and Accelerometer-Based Physical Variables in Small-Sided Games with and Without the Offside Rule. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, *236*(2), 134–142. https://doi.org/10.1177/1754337120987646

Dahlan, F., Hidayat, R., & Syahruddin, S. (2020). Pengaruh Komponen Fisik dan Motivasi Latihan Terhadap Keterampilan Bermain Sepakbola. *Jurnal Keolahragaan*, *8*(2), 126–139. https://doi.org/10.21831/jk.v8i2.32833

Dahlan, F., Samsudin, S., & Hernawan, H. (2024). The Influence of Small-Sided Conditioned Games (SSCG) and Interaction Patterns on Tactical Actions during Football Attacks. *Retos*, *59*, 75–83. https://doi.org/10.47197/retos.v59.106853

Eriksrud, O., & Westheim, F. (2025). Test-Retest and Intrasession Reliability of Assisted Sprint Outcome Measurements Using a Motorized Resistance Device. *Frontiers in Sports and Active Living*, *7*. https://doi.org/10.3389/fspor.2025.1592707

Federation Internationale de Football Association (FIFA). (2021). *Handbook for Technical Directors* (Federation Internationale de Football Association (FIFA), Ed.). Zurich, Switzerland: Federation Internationale de Football Association (FIFA).

Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decalogue for the Factor Analysis of Test Items. *Psicothema*, *34*(1), 7–17. https://doi.org/10.7334/psicothema2021.456

Folgado, H., Duarte, R., Marques, P., Gonçalves, B., & Sampaio, J. (2018). Exploring How Movement Synchronization is Related to Match Outcome in Elite Professional Football. *Science and Medicine in Football*, *2*(2), 101–107. https://doi.org/10.1080/24733938.2018.1431399

González-Víllora, S., Serra-Olivares, J., Pastor-Vicedo, J. C., & da Costa, I. T. (2015). Review of The Tactical Evaluation Tools for Youth Players, Assessing The Tactics in Team Sports: Football. *SpringerPlus*, *4*(1), 1–17. https://doi.org/10.1186/s40064-015-1462-0

Grgic, J., Lazinica, B., & Pedisic, Z. (2021). Test–Retest Reliability of The 30–15 Intermittent Fitness Test: A Systematic Review. *Journal of Sport and Health Science*, *10*(4), 413–418. https://doi.org/10.1016/j.jshs.2020.04.010

Guard, A., McMillan, K., & MacFarlane, N. (2022). Influence of Game Format and Team Strategy on Physical and Perceptual Intensity in Soccer Small-Sided Games. *International Journal of Sports Science and Coaching*, *17*(5), 1109–1118. https://doi.org/10.1177/17479541211056399

Hancock, A. B., Azul, D., Chennupati, S., Sediqi, N., Nyberg, J., & Nygren, U. (2025). Conceptual Model and Content Validity of the Voice and Communication Situation Questionnaire Developed With and for Transgender and Gender-Diverse People Presumed Female at Birth. *Journal of Speech, Language, and Hearing Research*, *68*(3), 963–986. https://doi.org/10.1044/2024_JSLHR-24-00203

Harper, L. D., Hunter, R., Parker, P., Goodall, S., Thomas, K., Howatson, G., … Russell, M. (2016). Test-Retest Reliability of Physiological and Performance Responses to 120 Minutes of Simulated Soccer Match Play. *Journal of Strength and Conditioning Research*, *30*(11), 3178–3186. https://doi.org/10.1519/JSC.0000000000001400

Kwon, T., Shin, S., & Shin, M. (2022). The Effect of Observational Learning on Self-Efficacy by Sport Competition Condition, Performance Level of Team Members, and Whether You Win or Lose. *International Journal of Environmental Research and Public Health*, *19*(16). https://doi.org/10.3390/ijerph191610148

Laakso, T., Davids, K., Luhtanen, P., Liukkonen, J., & Travassos, B. (2022). How Football Team Composition Constrains Emergent Individual and Collective Tactical Behaviours: Effects of Player Roles in Creating Different Landscapes for Shared Affordances in Small-Sided and Conditioned Games. *International Journal of Sports Science and Coaching*, *17*(2), 346–354. https://doi.org/10.1177/17479541211030076

Liveris, N. I., Tsarbou, C., Xergia, S. A., Papadopoulos, A., & Tsepis, E. (2024). Comparison of Inter-Rater and Intra-Rater Reliability of Raters with Different Levels of Experience When Using Landing Error Scoring System (LESS) in Field-Based Screening of Professional Football Players. *Sports*, *12*(9), 242. https://doi.org/10.3390/sports12090242

Ljunggren, G., Perera, N. K. P., & Hägglund, M. (2019). Inter-Rater Reliability in Assessing Exercise Fidelity for The Injury Prevention Exercise Programme Knee Control in Youth Football Players. *Sports Medicine - Open*, *5*(1), 35. https://doi.org/10.1186/s40798-019-0209-9

Milovanovic, M. (2023). Assessing The Reliability of Tests for Evaluation Technical Skills in Mini Basketball. *Journal of Physical Education and Sport*, *23*(11), 2917–2924. https://doi.org/10.7752/jpes.2023.11332

Muliyati, D., Septiningrum, A. D., Ambarwulan, D., & Astra, I. M. (2020). The Development of Guided Inquiry Student Worksheet using Tracker Video Analysis for Kinematics Motion Topics. *Journal of Physics: Conference Series*, *1491*(1), 1–7. https://doi.org/10.1088/1742-6596/1491/1/012062

Muñiz, J., & Fonseca-Pedrero, E. (2019). Ten Steps For Test Development. *Psicothema*, *31*(1), 7–16. https://doi.org/10.7334/psicothema2018.291

Narizuka, T., Takizawa, K., & Yamazaki, Y. (2023). Validation of a Motion Model for Soccer Players' Sprint by Means of Tracking Data. *Scientific Reports*, *13*(1), 865. https://doi.org/10.1038/s41598-023-27999-1

Nunes, N. A., Gonçalves, B., Coutinho, D., & Travassos, B. (2020). How Numerical Unbalance Constraints Physical and Tactical Individual Demands of Ball Possession Small-Sided Soccer Games. *Frontiers in Psychology*, *11*, 1464. https://doi.org/10.3389/fpsyg.2020.01464

Olthof, S. B. H., Frencken, W. G. P., & Lemmink, K. A. P. M. (2015). The Older, The Wider: On-Field Tactical Behavior of Elite-Standard Youth Soccer Players in Small-Sided Games. *Human Movement Science*, *41*, 92–102. https://doi.org/10.1016/j.humov.2015.02.004

Osborne, J. O., Kildalsen, I., Pedersen, S., Pettersen, S. A., Welde, B., Minahan, C. L., & Andersson, E. P. (2025). Test–Retest Reliability of Strength, Power, Agility, and Sprint Performance in Female Team Handball Players. *Journal of Science and Medicine in Sport*, *28*(3), 249–255. https://doi.org/10.1016/j.jsams.2024.11.008

Padrón-Cabo, A., Rey, E., Pérez-Ferreirós, A., & Kalén, A. (2019). Test–Retest Reliability of Skill Tests in the F-MARC Battery for Youth Soccer Players. *Perceptual and Motor Skills*, *126*(5), 1006–1023. https://doi.org/10.1177/0031512519866038

Práxedes, A., Del Villar Álvarez, F., Moreno, A., Gil-Arias, A., & Davids, K. (2019). Effects of a Nonlinear Pedagogy Intervention Programme on The Emergent Tactical Behaviours of Youth Footballers. *Physical Education and Sport Pedagogy*, *24*(4), 332–343. https://doi.org/10.1080/17408989.2019.1580689

Sequeira, C., Oliveira-Santos, M., Borges Rosa, J., Silva Marques, J., Oliveira Santos, E., Norte, G., & Gonçalves, L. (2024). Three-Dimensional Simulation for Interventional Cardiology Procedures: Face and Content Validity. *Revista Portuguesa de Cardiologia*, *43*(7), 389–396. https://doi.org/10.1016/j.repc.2023.11.006

Silva, P., Garganta, J., Araújo, D., Davids, K., & Aguiar, P. (2013). Shared Knowledge or Shared Affordances? Insights from an Ecological Dynamics Approach to Team Coordination in Sports. *Sports Medicine*, *43*(9), 765–772. https://doi.org/10.1007/s40279-013-0070-9

Slimani, M., Bragazzi, N. L., Tod, D., Dellal, A., Hue, O., Cheour, F., … Chamari, K. (2016). Do Cognitive Training Strategies Improve Motor and Positive Psychological Skills Development in Soccer Players? Insights from a Systematic Review. *Journal of Sports Sciences*, *34*(24), 2338–2349. https://doi.org/10.1080/02640414.2016.1254809

Snow, S., & Thomas, J. (2009). *Skills School Training Manual, Fundamental Ball Skills*. Frisco, Texas: US Youth Soccer Technical Department.

Teoldo da Costa, I., & de Freitas Silvino, M. P. (2023). Analysis of Tactical Behavior in Full- and Small-Sided Games: Comparing Professional and Youth Academy Athletes to Enhance Player Development in Soccer. *International Journal of Sports Science and Coaching*, *18*(1), 132–142. https://doi.org/10.1177/17479541221074992

Thai, T., Engel, L., Eades, S., Sanci, L., William, R., Gregory, P., … Mihalopoulos, C. (2025). Content and Face Validity of Preference-Based Quality of Life Instruments in Indigenous Youth Experiencing Mental Health Problems: A Qualitative Study. *Quality of Life Research*, *34*(4), 963–975. https://doi.org/10.1007/s11136-024-03870-2

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Expectional Children* (Special Ed; Minnesota: Leadership Training Institute, Ed.). Minneapolis: University of Minnesota.

Umar, Y., & Lian, L. H. (2024). Assessing Lecturers' Belief on Assessment in Nigeria Using Brown's Factor Model of Assessment Belief: A Study of Content-based Validity. *Pertanika Journal of Social Sciences and Humanities*, *32*(3), 1073–1093. https://doi.org/10.47836/pjssh.32.3.13

Yudiana, Y., Hidayat, Y., Hambali, B., & Slamat, S. (2017). Content Validity Estimation of Assessment Instrument Based on Volleyball Information System of Volleyball Learning: Field Research. *IOP Conference Series: Materials Science and Engineering*, Vol. 180. https://doi.org/10.1088/1757-899X/180/1/012230