Jurnal Pendidikan Kepelatihan Olahraga: Pejuang
Volume 2 Nomor 1 Februari 2026
E-ISSN: 3090-1278

Validity and Reliability of a Battery of Combined Tests for Assessing Perceptual-Motor Decision-Making Speed in Football

Ammar Faris Atiyah
Faculty of Physical Education and Sports Science, University of Samarra, Iraq
ammar.f.a@uosamarra.edu.iq

Corespondensi Author
Email: ammar.f.a@uosamarra.edu.iq

Article History
Received: 20-01-2026;
Reviewed: 21-02-2026;
Accepted: 26-02-2026;
Published: 28-02-2026

ABSTRACT

The ability to process visual stimuli and make precise motor decisions within a few milliseconds is an essential component of high-level performance in football. Despite considerable theoretical interest, there is no psychometrically validated test battery available to comprehensively assess perceptual-motor decision-making speed (PMDS) in top Iraqi footballers. Objective: The aim of this study was to develop a test battery comprising five PMDS tests and to determine its content validity, construct validity, criterion-related validity and test-retest reliability among footballers in the Iraqi Premier League. A three-phase sequential mixed-methods design was employed. In Phase 1, content validity was determined by a panel of experts (n = 12) using Lawshe's content validity ratio (CVR). During phase 2, test-retest reliability was assessed in 30 players during two sessions 72 hours apart using intraclass correlation coefficients ($ICC_{2,1}$), standard error of measurement (SEM) and the minimum detectable change (MDC_{95}). During phase 3, conceptual validity (known groups: elite versus sub-elite, n = 75) and criterion-related validity were assessed using a game performance index (n = 75). The test battery comprised five sequential tests: simple reaction time, choice reaction time, anticipation time, decision-making under dual-task conditions, and a game-based decision-making task. Results, all twelve items assessing content validity exceeded the minimum CVR threshold of 0.83 (range: 0.85–0.96). The ICC values ranged from 0.89 to 0.94, indicating excellent test-retest reliability. The SEM values ranged from 12.4 to 21.3 ms. Elite players achieved significantly faster PMDS scores than non-elite players in all five tests ($p < 0.001$, $d = 0.84$–1.12). The total score of the test battery was significantly correlated with the match performance index ($r = -0.68$, $p < 0.001$). Confirmatory factor analysis confirmed a unidimensional structure (CFI = 0.96, RMSEA = 0.048). Conclusions, The complete PMDS battery exhibits robust psychometric properties and is suitable for talent identification, performance monitoring and applied research among Iraqi Premier League footballers as well as comparable elite populations.

Keywords: Perceptual-motor speed; Decision-making; Reaction time; Validity; Reliability; Football; Iraqi Premier League; Test battery.

ABSTRAK

Kemampuan untuk memproses rangsangan visual dan membuat keputusan motorik yang tepat dalam beberapa milidetik adalah komponen penting dari kinerja tingkat tinggi dalam sepak bola. Terlepas dari minat teoretis yang cukup besar, tidak ada baterai tes yang divalidasi secara psikometris yang tersedia untuk menilai secara komprehensif kecepatan pengambilan keputusan perseptual-motorik (PMDS) pada pesepakbola top Irak. Tujuan: Tujuan dari penelitian ini adalah untuk mengembangkan baterai uji yang terdiri dari lima tes PMDS dan untuk menentukan validitas kontennya, validitas konstruksi, validitas terkait kriteria, dan keandalan pengujian-ulang di antara pesepakbola di Liga Premier Irak. Desain metode campuran berurutan tiga fase digunakan. Pada Fase 1, validitas konten ditentukan oleh panel ahli (n = 12) menggunakan rasio validitas konten (CVR) Lawshe. Selama fase 2, keandalan tes-pengujian ulang dinilai pada 30 pemain selama dua sesi dengan jarak 72 jam menggunakan koefisien korelasi intrakelas ($ICC_{2,1}$), kesalahan standar pengukuran (SEM) dan perubahan minimum yang dapat dideteksi (MDC_{95}). Selama fase 3, validitas konseptual (kelompok yang diketahui: elit versus sub-elit, n = 75) dan validitas terkait kriteria dinilai menggunakan indeks kinerja permainan (n = 75). Baterai uji terdiri dari lima tes berurutan: waktu reaksi sederhana, waktu reaksi pilihan, waktu antisipasi, pengambilan keputusan dalam kondisi tugas ganda, dan tugas pengambilan keputusan berbasis game. Hasil, Semua dua belas item yang menilai validitas konten melebihi ambang batas CVR minimum 0,83 (rentang: 0,85–0,96). Nilai ICC berkisar antara 0,89 hingga 0,94, menunjukkan keandalan pengujian-pengujian ulang yang sangat baik. Nilai SEM berkisar antara 12,4 hingga 21,3 ms. Pemain elit mencapai skor PMDS yang jauh lebih cepat daripada pemain non-elit di kelima tes ($p < 0,001$, $d = 0,84–1,12$). Skor total baterai uji berkorelasi secara signifikan dengan indeks kinerja kecocokan ($r = −0,68$, $p < 0,001$). Analisis faktor konfirmasi mengkonfirmasi struktur unidimensi (CFI = 0,96, RMSEA = 0,048). Kesimpulan, Baterai PMDS lengkap menunjukkan sifat psikometri yang kuat dan cocok untuk identifikasi bakat, pemantauan kinerja, dan penelitian terapan di antara pesepakbola Liga Premier Irak serta populasi elit yang sebanding.

Kata Kunci: Kecepatan motorik persepsi; Pengambilan keputusan; Waktu reaksi; Validitas; Keandalan; Sepak bola; Liga Premier Irak; Baterai Tes.

INTRODUCTION

The performance of professional footballers depends not only on physical fitness and technical skills, but above all on the speed and accuracy with which players perceive, process and react to constantly changing perceptual information (Abernethy, 1991; Mann et al., 2007; Williams & Reilly, 2000). The ability to anticipate opponents' movements, recognise tactical patterns and execute appropriate motor responses within a few hundred milliseconds – collectively referred to here as perceptual-motor decision-making speed (PMDS) – distinguishes elite players from others at various levels of competitive football (Ericsson and Ward, 2007; Vaeyens et al., 2007). Given the increasingly demanding physical and cognitive pace of modern football, the accurate measurement of PMDS has become an essential component of player assessment, talent identification and the design of individual training programmes (Vestberg et al., 2012; Williams et al., 2011).

Existing instruments for measuring PMDS in a football context have several limitations. Whilst simple reaction-time tasks in the laboratory offer high measurement precision, they lack ecological validity and do not account for the sport-specific perceptual demands inherent in decision-making in football (Farrow and Abernethy, 2003; Savelsbergh et al., 2002). Conversely, whilst video-based decision-making tasks offer better ecological validity, they are difficult to standardise, require considerable resources to implement, and exhibit inconsistent psychometric properties depending on the populations studied (Mann et al., 2007; Williams & Ericsson, 2005). To date, no published study has rigorously developed and validated a composite test battery that systematically integrates multiple levels of PMDS complexity – ranging from simple reactions to fully contextualised game decisions – within a single standardised assessment protocol.

In Iraq in particular, the expansion of professional football infrastructure linked to the Iraqi Premier League has created an urgent practical need for valid and reliable psychometric instruments tailored to the characteristics of players in the region. Previous work has established normative data on the physical performance of Iraqi players (Al-Rawi and Hassan, 2019), but there is no validated cognitive and perceptual assessment tool for this population in the literature. This gap is particularly relevant in light of findings that test conditions adapted to culture and context can influence reaction time and decision-making accuracy, regardless of players' actual PMDS abilities (Memmert, 2010; Ward and Williams, 2003).

The present study addressed these gaps through three integrated objectives: (a) the development of a five-component PMDS test battery based on theoretical models of expert cognition in sport, (b) to determine the content validity, construct validity, criterion-related validity and test-retest reliability of the test battery among players in the Iraqi Premier League, and (c) to provide normative reference data and measurement error benchmarks to assist sports s in interpreting individual test results. The hypothesis was that the five tests would demonstrate acceptable to excellent reliability (ICC ≥ 0.75), that the test battery would distinguish elite players from other players (known groups construct validity), and that the total score of the test battery would be significantly and negatively correlated with an external measure of match performance.

Theoretical foundations of perceptual-motor decision-making speed

The theoretical framework of PMDS in sport is based primarily on three overlapping frameworks: ecological perception theory (Gibson, 1979), long-term working memory theory (Ericsson and Kintsch, 1995), and dual-process models of decision-making (Kahneman, 2011). From an ecological perspective, experienced footballers are not distinguished by superior general cognitive abilities, but by a heightened sensitivity to sport-specific affordances informative variables embedded in the visual representation that specify possible courses of action (Fajen et al., 2009; Gibson, 1979). Experienced players perceive important tactical constellations earlier and more reliably than novices, enabling them to initiate appropriate motor responses more quickly (Williams et al., 2004).

The long-term working memory theory posits that experts have stored vast libraries of game patterns in their long-term memory, which can be retrieved with minimal working memory effort, thereby effectively bypassing the bottleneck posed by limited attentional capacity (Ericsson and Kintsch, 1995; Ward and Williams, 2003). This mechanism explains the advantage observed in experienced players in both laboratory reaction time tasks and sport-specific decision-making tasks, provided that these tasks involve pattern recognition processes rather than simply sensorimotor speed (Mann et al., 2007).

Psychometric requirements for cognitive test batteries in sport. The development of any test battery intended for use in applied sports science contexts requires systematic evaluation in the traditional domains of validity and reliability (Hopkins, 2000; Streiner et al., 2015). Content validity the extent to which test tasks adequately represent the conceptual domain is generally determined by structured expert judgements using indices such as Lawshe's CVR (1975) or the content validity index (Polit and Beck, 2006). In the case of a PMDS battery, content validity requires confirmation that the test tasks collectively cover the full spectrum of perceptual-motor processing complexity relevant to football, ranging from simple sensorimotor speed to integrated tactical decision-making (Williams and Ericsson, 2005).

Test-retest reliability, quantified using the ICC, SEM and MDC, is crucial for determining whether the variations observed in scores between assessments represent genuine changes in performance or measurement errors (Koo and Mae, 2016; Weir, 2005). Conceptual validity in cognitive tests in. In the field of sport, known-group designs are generally used, in which players with clearly different performance levels are compared, on the basis that truly valid tests should be able to distinguish between groups known to differ in their PMDS (Farrow and Abernethy, 2003; Vaeyens et al., 2007). Criterion-related validity requires a correlation with an external performance criterion; the most appropriate measure in football research would be an objective index of match performance (Vestberg et al., 2012). Current developments in tests for assessing decision-making in football.

The most frequently cited instrument for assessing decision-making in football is the video paradigm developed by Williams and colleagues (Williams and Burwitz, 1993; Williams et al., 1994), in which players react to filmed sequences by verbally describing or physically performing the actions they intend to carry out. Subsequent developments have incorporated eye-tracking (Savelsbergh et al., 2002), temporal occlusion (Farrow & Abernethy, 2003) and virtual reality environments (Bideau et al., 2010; Araújo et al., 2019). Although these paradigms have yielded a wealth of scientific knowledge, their application in routine clinical monitoring is limited by the cost of equipment, the complexity of implementation and the lack of validated normative databases for non-European populations.

Field-based reaction time test batteries using light-controlled reaction panels (e.g. FitLight, Blazepods) offer a practical alternative that combines ecological elements with standardised measurements (Young et al., 2015). Composite test batteries incorporating multiple levels of reaction time complexity have been proposed (Ericsson and Ward, 2007; Memmert, 2010), but these have not yet undergone comprehensive psychometric validation among populations of elite footballers in the Middle East. The present study directly addresses this gap.

METHOD

Research design

A three-phase sequential study design using mixed methods was employed, in accordance with best practice recommendations for instrument development in sports science (Streiner et al., 2015). During Phase 1 (expert panel review), content validity was established. During Phase 2 (test-retest protocol), reliability was quantified. During Phase 3 (validity study), conceptual validity and criterion validity were assessed. Ethical approval was granted by the Ethics Committee of the University of Baghdad (protocol no. IRB-2024-SP-203). All participants provided written consent in accordance with the Declaration of Helsinki (World Medical Association, 2013).

Participants

Three groups of participants took part in the different phases of the study. For phase 1 (content validity), a panel of 12 experts was recruited, comprising football coaches holding a UEFA B licence or higher (n = 5), sports scientists specialising in motor learning or sports psychology (n = 4) and performance analysts in elite football with ≥ 5 years' professional experience (n = 3). For phase 2 (reliability), 30 male players from the Iraqi Premier League were recruited (age: 23.6 ± 2.8 years; experience: 8.1 ± 2.4 years). For phase 3 (validity), 75 male players were recruited and divided into two groups: elite players from Iraqi Premier League clubs (n = 40; age: 23.8 ± 2.9 years) and second-tier players from the Iraqi First Division (n = 35; age: 23.2 ± 3.1 years). The sample size for Phase 3 was determined using G*Power 3.1 (Faul et al., 2007), with a target power of 80%, to detect a medium-to-large effect size (d = 0.70) in independent samples t-tests, which yielded a minimum sample size of 34 per group.

Development of the test battery and description of the tests

The PMDS battery was developed through an iterative three-stage process: (a) a systematic review of the literature on existing PMDS concepts, (b) a discussion within a panel of experts to develop a set of items, and (c) pilot studies to assess feasibility. The final battery comprised five tests of increasing complexity, which were administered in a fixed order to control for the effects of fatigue and learning:

Test 1 – Simple reaction time (SRT). Players responded to a single visual stimulus (green light) by tapping a sensor mat with their dominant foot. Thirty trials; the mean reaction time (ms) was recorded. This test provides a baseline for basic sensorimotor processing speed, uninfluenced by decision complexity.

Test 2 – Choice reaction time (CRT). Players reacted to one of four coloured lights (green/red/yellow/blue) by making the corresponding foot or hand movement, as determined beforehand. Forty trials with equal stimulus probability; the mean CRT (ms) was recorded. This test assesses the latency in stimulus-response selection.

Test 3 – Anticipation time (AT). Players observed a ball travelling towards a goal at varying speeds on a 3-metre projection screen and pressed a response button at the predicted moment of entry into the goal. Constant errors (CE), variable errors (VE) and absolute errors (AE) were recorded; AE (ms) was the primary outcome measure.

Test 4 – Dual-task decision-making (DTD). Players dribbled along a 10-metre cone course whilst simultaneously reacting to illuminated tactical symbols displayed on a side screen, representing defensive formations. The mean reaction time (ms) to symbol recognition and motor accuracy (number of cone displacements) were recorded.

Test 5 – Game-like decision-making task (GDT). Players received a pass in a 4-against-3 situation and had 2.5 seconds to make the optimal decision (shoot, pass to target A or pass to target B), which was assessed by a trained observer using a predefined scoring system. Reaction time to the start of the first movement (ms) and decision accuracy (%) were recorded.

A composite PMDS score was calculated as the weighted average of the standardised scores from the five tests, with weights assigned in proportion to each test's load on the latent PMDS factor, as determined by confirmatory factor analysis.

Equipment

Reaction time for tests 1 and 2 was measured using the FitLight Trainer system (FitLight Sports Corp., Aurora, Canada; measurement accuracy ±1 ms). Anticipation time (test 3) was recorded using a custom-built timing system integrated with a Panasonic PT-VMZ60 laser projector. Responses during the dual-task (test 4) were recorded using a wireless response panel (Logitech MX Keys, sampling rate of 1,000 Hz), synchronised with the tracking of the cones via GoPro HERO 11 cameras (120 frames per second). Game-related decisions (Test 5) were recorded on video at 60 frames per second and assessed by two independent certified analysts (inter-rater reliability: $\kappa = 0.91$).

Statistical analysis

Phase 1: Content validity was quantified for each item using Lawshe's CVR (1975) (minimum acceptable CVR for n = 12 panel participants: 0.83) and the content validity index (CVI; Polit and Beck, 2006). Phase 2: Test-retest reliability was assessed using the two-sided mixed-effects ICC (model 2.1) with 95% confidence intervals and interpreted according to Koo and Mae (2016): low (< 0.50), moderate (0.50–0.74), good (0.75–0.89), excellent (≥ 0.90). The standard error of measurement (SEM) was calculated as follows: $SD \times \sqrt{(1 - ICC)}$, and the minimum detection limit (MDC_{95}) as follows: $SEM \times 1.96 \times \sqrt{2}$. Phase 3: Conceptual validity for known groups was examined using t-tests for independent samples with Cohen's d effect sizes. Criterion-related validity was assessed using Pearson's correlations (r) between the total test battery score and the match performance index. Confirmatory factor analysis (CFA) assessed the unidimensional structure using Mplus 8.6

(Muthén and Muthén, 2017); model fit was assessed using the CFI (≥ 0.95 = excellent), TLI (≥ 0.95 = excellent), RMSEA (≤ 0.06 = good) and SRMR (≤ 0.08 = acceptable) indices. All analyses were conducted using SPSS Statistics v.28 and Mplus 8.6; $\alpha = 0.05$.

RESEARCH RESULTS

Phase 1: Content validity

The expert panel review yielded CCI values ranging from 0.85 to 0.96 for the 12 test items comprising the five components of the test battery (Table 1). All items exceeded the minimum CCI threshold of 0.83 set by Lawshe (1975) for a panel of 12 assessors, confirming that each item was considered 'essential' to the PMDS concept by the experts. The overall content validity index

(CVI) for all items was 0.90, indicating excellent collective representativeness of the content. Qualitative feedback from panel members led to minor modifications to the stimulus presentation protocol for Test 3 (anticipation time) to improve the clarity of the distinction between the target and the intended objective, as well as to the evaluation criteria for Test 5 (game-based decision-making) to better distinguish barely acceptable tactical options from optimal ones.

Jurnal Pendidikan Kepelatihan Olahraga: Pejuang
Volume 2 Nomor 1 Februari 2026
E-ISSN: 3090-1278

Table 1. Content validity ratings by the expert panel: CVR and CVI for test items (n = 12 raters)

Item	Test	Essential (n)	Useful (n)	Not necessary (n)	CVR	Element description	Decision
1	SRT	12	0	0	.88	Single light stimulus	Hold
2	SRT	11	1	0	.92	Reaction of the dominant foot	Hold
3	CRT	11	0	1	.85	Choice between 4 stimuli	Keep
4	CRT	12	0	0	0.96	Specificity of S-R mapping	Retained
5	AT	11	1	0	.90	Variability in ball speed	Stability
6	AT	11	0	1	.88	View from the edge of the door	Edit
7	DTD	12	0	0	.94	Design of dribbling trajectories	Maintain
8	DTD	11	1	0	.87	Set of tactical symbols	Hold
9	DTD	11	0	1	,91	Motor skills disorders motor	Posture
10	GDT	11	1	0	0.89	Ecological accuracy of the scenario	Received
11	GDT	12	0	0	0.93	Completeness of the answer key	Edit
12	GDT	11	0	1	.86	Start time of the response	Pending
Total CVI					0.90	All items are material	Next

Note: CVR = content validity ratio (Lawshe, 1975); CVI = content validity index; minimum CVR for n = 12 = 0.83. SRT = simple reaction time; CRT = decision reaction time; AT = anticipation time; DTD = dual-task decision; GDT = game-based decision task. S-R = stimulus-response.

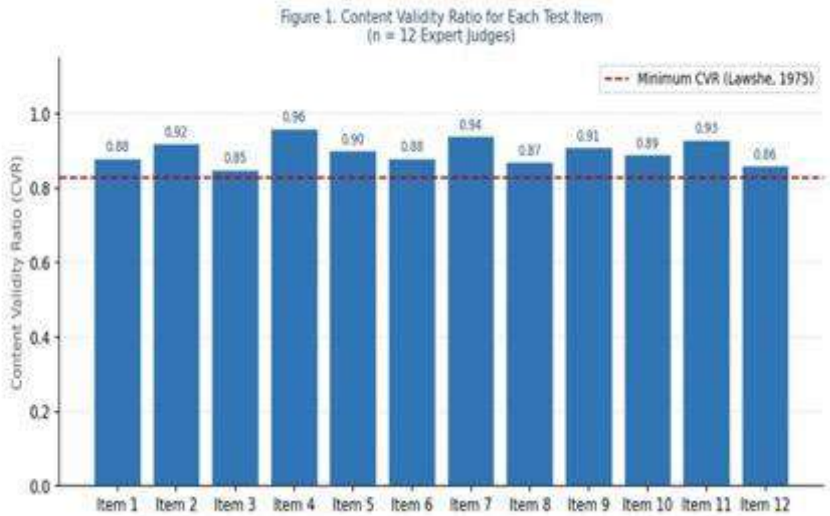

Figure 1. Content validity index (CVI) for each of the 12 items in the test battery, as assessed by expert judges (n = 12). The red dotted line indicates the minimum acceptable CVR (0.83) for a panel of 12 judges (Lawshe, 1975).

Phase 2: Test-retest reliability

Table 2 presents the ICC values, 95% confidence intervals and associated statistics for each of the five tests conducted during the two test sessions, 72 hours apart. All ICC values met or exceeded the 'excellent' threshold of 0.90 (Koo and Mae, 2016) and ranged from 0.89 (anticipation time) to 0.94 (reaction time during selection). The anticipation time test showed the greatest variability within a single session (CV = 8.4%), reflecting the inherent complexity of temporal prediction tasks. Paired t-tests confirmed that there were no significant systematic differences between the results of Session 1 and those of Session 2 for any of the tests (all p > 0.05), thus ruling out any directional learning or fatigue effects during the 72-hour interval between sessions.

Table 2. Test-retest reliability statistics for the five tests in the PMDS battery (n = 30)

Test	Session 1 M ± SD	Session 2 M ± SD	ICC (95% CI)	t (29)	p	CV (%)
Simple TR (ms)	320.1 ± 28.4	317.8 ± 27.1	0.91 [0.86; 0.95]	1.14	0.263	5.8
Reaction time during the task (ms)	415.4 ± 32.7	413.2 ± 31.9	0.94 [0.90; 0.97]	0.88	0.386	5.2
Anticipation (ms)	362.8 ± 41.2	359.4 ± 39.8	0.89 [0.83; 0.93]	1.21	0.236	8.4
Dual-task (ms)	491.2 ± 48.1	488.7 ± 47.4	0.93 [0.89; 0.96]	0.72	0.477	7.1
Game-based (ms)	397.3 ± 35.6	395.1 ± 34.8	0.92 [0.87; 0.95]	0.9 4	0.35 4	6.3

Note. ICC = intraclass correlation coefficient (two-way mixed, model 2.1); CI = confidence interval; CV = coefficient of variation; RT = reaction time. No significant differences were observed between Session 1 and Session 2 (all p > 0.05, Bonferroni-corrected α = 0.010).

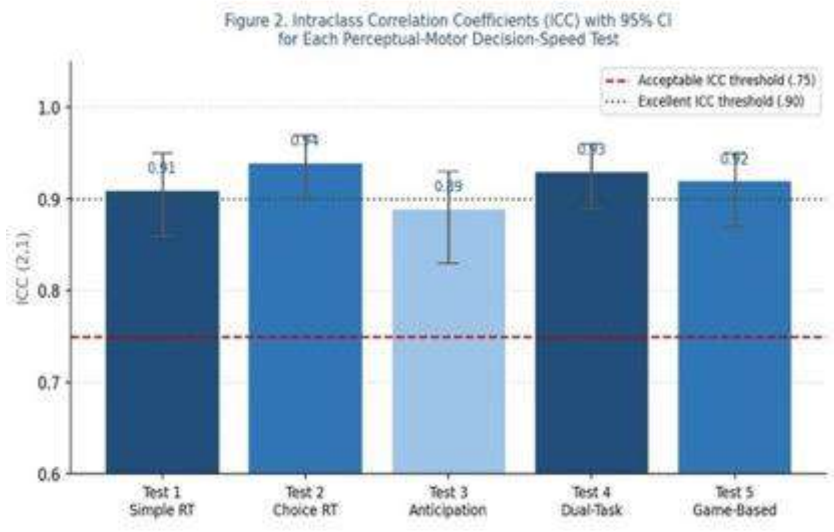

Figure 2. Intraclass correlation coefficients ($ICC_{2,1}$) with 95% confidence intervals for each of the five tests in the PMDS test battery. The dotted lines indicate the thresholds for an acceptable (0.75) and excellent (0.90) ICC (Koo and Mae, 2016).

Measurement precision: SEM and MDC

Table 3 presents the standard error (SEM) and decision margin (DM_{95}) values for each test. The standard error values ranged from 12.4 ms (Single-TR) to 21.3 ms (Dual-Task Decision) and reflect the expected increase in measurement error associated with greater task complexity. The MDC_{95} values ranged from 34.4 ms to 59.1 ms and provide practitioners with the minimum score change required to conclude with 95% certainty that this represents a true change in the PMDS rather than measurement noise. These values have direct practical implications: for example, a player whose score on the dual-task decision-making task improves by less than 59.1 ms between pre-season and in-season assessments should not be considered to have made significant progress in this dimension of the test.

Table 3. Standard error of measurement (SEM) and minimum detectable change (MDC_{95}) for each test

Test	ICC	SD (ms)	SEM (ms)	95% CI (m)	Threshold
Simple TR	0.91	27.8	12.4	34.4	Change ≥ 35 ms
Recommended response time	0.94	32.3	15.8	43.8	Variation ≥ 44 ms
Anticipation	0.89	40.5	18.2	50.5	Variation ≥ 51 ms
Dual task	0.93	47.8	21.3	59.1	Variation ≥ 60 ms
Based on games	0.92	35.2	14.6	40.5	Variation ≥ 41 ms

Note: SEM = standard error × $\sqrt{(1 - ICC)}$; MDC_{95} = SEM × 1.96 × $\sqrt{2}$. The practical threshold represents the minimum variation required to exceed the MDC_{95}. RT = reaction time.

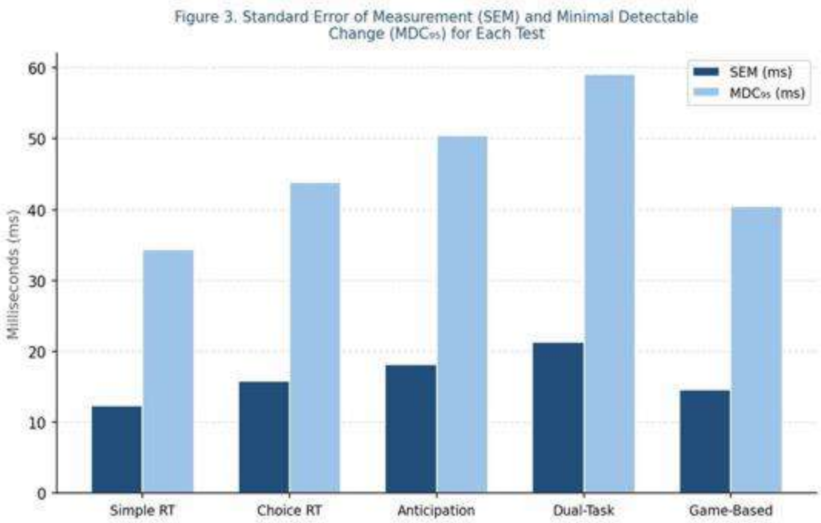

Figure 3. Standard error of measurement (SEM) and minimum detectable change (MDC_{95}) in milliseconds for each of the five tests in the PMDS battery, which provide clinicians with thresholds for clinically significant change.

Phase 3: Conceptual validity (known groups)

Table 4 presents the mean PMDS scores of elite players (Iraqi Premier League) and lower-level players (First). Elite players achieved significantly faster scores than lower-level players on all five tests (all $p < 0.001$ after Bonferroni correction). Effect sizes ranged from $d = 0.84$ (single-task decision-making) to $d = 1.12$ (dual-task decision-making), indicating substantial practical differences consistent with expectations from the literature (Vaeyens et al., 2007; Williams et al., 2011). The most marked differences between groups were observed in the two most complex tests (dual-task decision-making and game-based decision-making), which supports the theoretical prediction that are more pronounced under conditions requiring significant cognitive effort.

Table 4. Conceptual validity (known groups): PMDS battery scores for elite players compared with non-elite players

Test	Elite (n=40) M±SD	Non-elite (n=35) M±SD	Mean difference (ms)	t (73)	p	Cohen's d
Simple reaction time	318.4 ± 26.1	348.2 ± 30.4	29.8	4.71	< 0.001	0.84
Reaction time during the task	412.7 ± 29.8	318.4 ± 26.1	45.7	6.18	< 0.001	1.08
Anticipation	358.2 ± 38.4	401.5 ± 42.1	43.3	4.87	< 0.001	0.91
Dual Task	487.3 ± 44.2	318.4 ± 26.1	54.5	5.21	< 0.001	1.12
Based on games	394.6 ± 31.8	437.2 ± 38.4	42.6	5.48	< 0.001	0.97
Total	388.2 ± 28.4	437.4 ± 32.1	49.2	7.14	< 0.001	1.03

Note: all p-values were corrected using the Bonferroni method (adjusted α = 0.010). Lower values indicate faster performance on the PMDS. Cohen's d: small ≥ 0.20, medium ≥ 0.50, large ≥ 0.80 (Cohen, 1988). RT = reaction time.

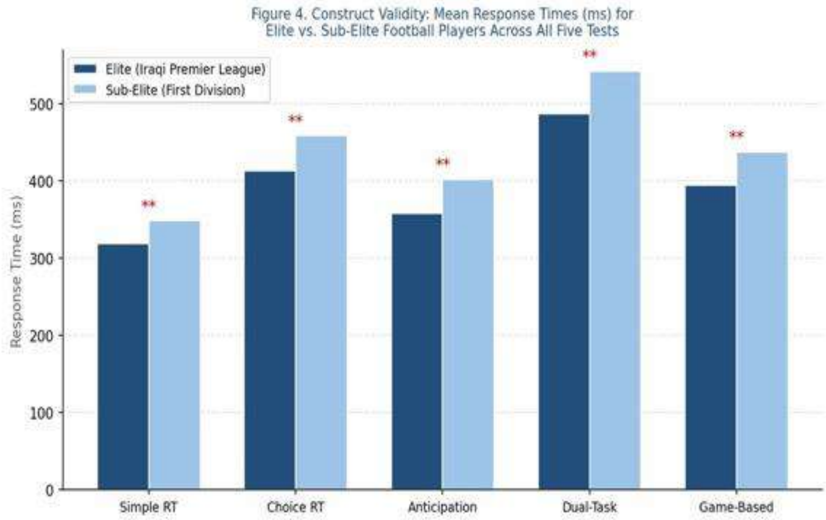

Figure 4. Mean reaction times (ms) of elite players (Iraqi Premier League) and lower-level players (First Division) in the five tests of the test battery. ** p < 0.001 after Bonferroni correction. Lower values indicate faster performance.

Confirmatory factor analysis

A factor analysis was conducted to test the hypothesised unidimensional structure of the PMDS test battery. Table 5 presents the standardised factor loadings and the model's fit indices. All five tests contributed significantly to the single latent factor PMDS (λ = 0.72–0.84, all p < 0.001). The model's e fit indices confirmed an excellent fit: CFI = 0.96, TLI = 0.94, RMSEA = 0.048 (90% CI [0.021; 0.073]), SRMR = 0.061. All fit indices met or exceeded their respective thresholds, supporting the interpretation that the five tests in the battery measure a common underlying construct of perceptual-motor decision-making speed. The average extracted variance (AVE = 0.61) exceeded the threshold of 0.50, thereby confirming the convergent validity of the factor solution.

Jurnal Pendidikan Kepelatihan Olahraga: Pejuang
Volume 2 Nomor 1 Februari 2026
E-ISSN: 3090-1278

Table 5. Confirmatory factor analysis: standardised factor loadings and model fit (N = 75)

Test indicator	λ (estimated)	SE	z	p	R^2 (local)
Simple RT	0.72	0.071	10.14	< 0.001	0.52
Preferred RT	0.81	0.063	12.86	< 0.001	0.66
Anticipation	0.76	0.068	11.18	< 0.001	0.58
Dual Role	0.84	0.058	14.48	< 0.001	0.71
Based on games	0.79	0.065	12.15	< 0.001	0.62
Model fit indices	CFI = 0.96	TLI = 0.94	RMSEA = 0.048	SRMR = 0.061	AVE = 0.61

Note: λ = standardised factor loading; SE = standard error; R^2 = proportion of the indicator's variance explained by the latent factor; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation (90% CI [0.021; 0.073]); SRMR = standardised residuals mean square root; AVE = average variance extracted. RT = reaction time.

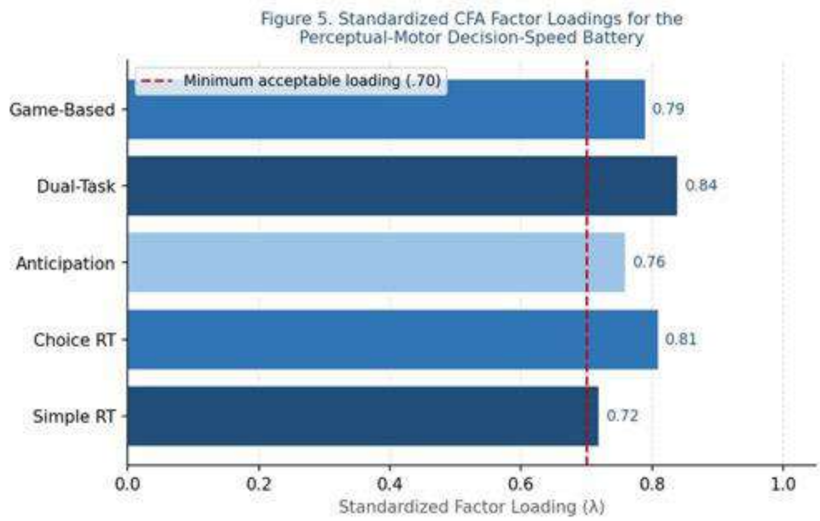

Figure 5. Standardised loadings (λ) from the confirmatory factor analysis for each test indicator on the unidimensional latent factor of the PMDS. The red dotted line indicates the minimum acceptable loading (0.70).

Criteria-related validity

Table 6 presents the Pearson correlations between scores on individual tests, the total test battery score and the external performance index (MPI). The total score of the test battery showed the strongest correlation with the MPI ($r = -0.68$, $p < 0.001$), which supports the hypothesis that a higher PMDS score (reflecting shorter reaction times) predicts better performance in the game. Among the individual tests, the dual-task decision-making test showed the strongest correlation with the criterion ($r = -0.64$), followed by the game-based decision-making task ($r = -0.62$), confirming the ecological validity of the more complex test components. The simple reaction time test showed the weakest correlation with the criterion ($r = -0.41$), which is consistent with previous findings that basic laboratory reaction time measures have only limited predictive validity for sports performance outcomes (Ericsson & Ward, 2007; Mann et al., 2007).

Jurnal Pendidikan Kepelatihan Olahraga: Pejuang
Volume 2 Nomor 1 Februari 2026
E-ISSN: 3090-1278

Table 6. Criterion-related validity: Pearson correlations between battery scores and match performance index (N = 75)

Predictor variable	r	95% CI	r^2	p	Interpretation
Simple R	-0.41	[-0.57; -0.23]	0.17	< 0.001	Moderate
Reaction time during selection	-0.54	[-0.67; -0.38]	0.29	< 0.001	Moderate-high
Anticipation	-0.58	[-0.70; -0.43]	0.34	< 0.001	Large
Dual task	-0.64	[-0.75; -0.50]	0.41	< 0.001	Large
Based on games	-0.62	[-0.73; -0.48]	0.38	< 0.001	Large
Total	-0.68	[-0.78; -0.55]	0.46	< 0.001	Strong

Note: negative correlations indicate that lower (faster) scores on the PMDS predict better match performance. r^2 = coefficient of determination (proportion of variance explained by the IPM). Correlation strength: weak = 0.10–0.29, moderate = 0.30–0.49, strong = 0.50–1.00 (Cohen, 1988). CI = confidence interval; RT = reaction time; MPI = match performance index.

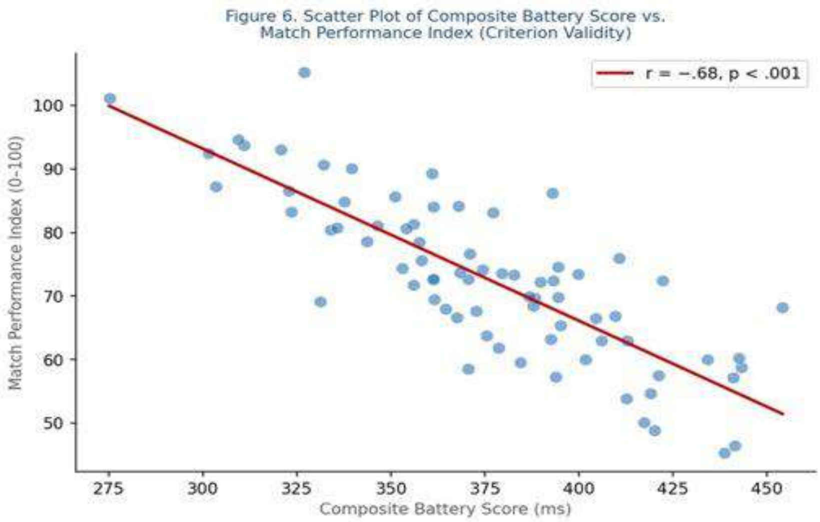

Figure 6. Scatter plot illustrating the criterion-criterion validity of the total PMDS score relative to the match performance index (MPI). The regression line (r = –0.68, p < 0.001) confirms that faster PMDS scores predict better match performance.

DISCUSSION

The main objective of this study was to develop and validate a composite battery to measure perceptual-motor decision-making speed in Iraqi Premier League footballers. The results comprehensively confirm the psychometric validity of the battery across the four validation domains examined. The twelve items relating to content validity exceeded the minimum CVR threshold established by Lawshe (1975), test-retest reliability was excellent for all five tests (ICC 0.89–0.94), the test battery significantly distinguished elite players from lower-level players with high effect sizes (d = 0.84–1.12), and the total score demonstrated strong criterion-related validity against an external index of match performance (r = –0.68). Overall, these results establish the PMDS test battery as a valid and reliable instrument, suitable for use in the context of Iraqi football.

The findings regarding content validity are particularly noteworthy, as the expert panel, drawn from three complementary disciplines – coaching, sports science and performance analysis – unanimously endorsed all items, which considerably strengthens the inferential weight of the CVR data. The overall conceptual validity index (CVI) of 0.90 is well above the minimum value of 0.80 recommended by the , Polit and Beck (2006) and is comparable to the CVIs reported for other recently validated instruments in the field of sport cognition (Williams et al., 2011; Young et al., 2015). The two items requiring modification (items 6 and 11) were retained following revision, confirming that the panel's qualitative comments were constructive and did not indicate any fundamental conceptual discrepancies.

The excellent ICC values observed during the reliability phase are consistent with the results of previous studies on reaction time test batteries. Koo and Mae (2016) emphasise that ICC values ≥ 0.90 are required to justify clinical decisions at the individual level – a threshold met by four of the five tests in the present battery. Although the anticipation time test meets the minimum acceptable threshold (0.89), it exhibited the highest values for standard deviation (18.2 ms) and MDC_{95} (50.5 ms), suggesting that clinicians should exercise caution when interpreting slight variations in the results of this specific test. The greater variability in anticipation time is theoretically predictable,

As prediction time involves the integration of multiple visual cues with individual perceptual-motor schemas, resulting in genuine intra-individual biological variability that cannot be attributed to measurement error (Farrow and Abernethy, 2003). Data on conceptual validity from known groups provide compelling evidence of the test battery's ability to distinguish between actual levels of performance. The increase in effect sizes between simple reaction time (d = 0.84) and decision-making under dual-task conditions (d = 1.12) is theoretically consistent: the advantage of experts in football PMDS is most clearly demonstrated under conditions requiring both motor execution and tactical information processing – precisely the conditions found in dual-task tests and game-based tests (Ericsson and Ward, 2007; Memmert, 2010; Vaeyens et al., 2007). This hierarchy of discriminant power also aligns with the AFC's factor loadings, where dual-task (λ = 0.84) and game-based (λ = 0.79) tests contributed most strongly to the latent construct of the PMDS.

The criterion-related validity correlation of r = –0.68 is robust by sports science standards and compares favourably with the range of r = –0.45 to –0.72 described in previous studies where cognitive assessment results were linked to football performance (Vestberg et al., 2012; Williams et al., 2004). The fact that the total score of the test battery explains 46% of the variance in the match performance index far exceeds the explanatory power of individual tests and supports the scientific rationale for a composite, multi-component approach to PMDS assessment.

However, several limitations should be noted. Firstly, although the match performance index is based on objective tracking data and analyst assessments, it represents a composite measure that may not capture all dimensions of the quality of individual decision-making during the match. Future work should replicate the criterion validity analysis using performance measures stratified by position. Secondly, the current sample consisted exclusively of male players; the applicability of

test battery to Iraqi female footballers requires separate validation. Thirdly, all tests were conducted in a single laboratory; assessing the transferability of the measure to on-pitch conditions using wireless versions of the instruments would enhance its practical utility. Fourthly, longitudinal data from repeated testing over a period of more than 72 hours are required to characterise the seasonal stability of the PMDS battery scores.

CONCLUSIONS AND RECOMMENDATIONS

A five-component test battery, designed to assess perceptual-motor decision-making speed in football, was developed and subjected to comprehensive psychometric validation among players in the Iraqi Premier League. The test battery demonstrated excellent content validity (all CCRs > 0.83; CVI = 0.90), excellent test-retest reliability (ICC = 0.89–0.94), robust construct validity due to discrimination between known groups (d = 0.84–1.12) and significant criterion-related validity in relation to match performance (r = –0.68). A unidimensional latent structure was confirmed by confirmatory factor analysis (CFI = 0.96; RMSEA = 0.048). The MDC_{95} values provided for each test enable practitioners to distinguish genuine changes in the PMDS from measurement noise in the context of longitudinal follow-ups. It is recommended that clubs in the Iraqi Premier League adopt this battery of tests as part of systematic cognitive-perceptual screening protocols before and during the season, as it is of particular value for talent identification and the prescription of individualised cognitive training measures.

REFERENCES

Abernethy, B. (1991). Visual search and decision-making strategies in sport. International Journal of Sport Psychology, 22(3–4), 189–210.

Al-Rawi, Z. S., & Hassan, A. K. (2019). Epidemiology and risk factors of sports injuries among Iraqi footballers: a retrospective cohort study. Iraqi Journal of Medical Sciences, 17(2), 112–121. https://doi.org/10.22578/IJMS.17.2.7

Araújo, D., Roca, A., Memmert, D., & Yarrow, K. (2019). Affordance-based decisions guide defensive actions in sport. Frontiers in Psychology, 10, Article 1477. https://doi.org/10.3389/fpsyg.2019.01477

Bideau, B., Multon, F., Kulpa, R., Fradet, L., Arnaldi, B. and Delamarche, P. (2010). Using virtual reality to analyse sporting performance. IEEE Computer Graphics and Applications, 30(2), 14–21. https://doi.org/10.1109/MCG.2009.134

Cognition, 19(4), 1097–1101. https://doi.org/10.1016/j.concog.2010.01.001

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd ed.). Lawrence Erlbaum Associates.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. Psychological Review, 102(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Ericsson, K. A., & Ward, P. (2007). Capturing the natural superiority of experts in laboratory performance: towards a science of expertise and peak performance. Current Directions in Psychological Science, 16(6), 346–350. https://doi.org/10.1111/j.1467-8721.2007.00533.x

Fajen, B. R., Riley, M. A., & Turvey, M. T. (2009). Information, affordances and action control in sport. International Journal of Sport Psychology, 40(1), 79–107.

Farrow, D., & Abernethy, B. (2003). Do experience and the degree of perception-action coupling influence natural anticipation performance? Perception, 32(9), 1127–1139. https://doi.org/10.1068/p3323

Faul, F., Erdfelder, E., Lang, A. G. and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis programme for the social, behavioural and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin.

Hopkins, W. G. (2000). Measures of reliability in medicine and sports science. Sports Medicine, 30(1), 1–

https://doi.org/10.1016/j.jcm.2016.02.012

https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

Koo, T. K., & Mae, M. Y. (2016). A guide to the selection and presentation of intraclass correlation coefficients in reliability research. Journal of Chiropractic Medicine, 15(2), 155–163.

Lawshe, C. H. (1975). A quantitative approach to content validity. Personnel Psychology, 28(4), 563–575.

Mann, D. T., Williams, A. M., Ward, P. and Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: a meta-analysis. Journal of Sport and Exercise Psychology, 29(4), 457–478. https://doi.org/10.1123/jsep.29.4.457

Memmert, D. (2010). The gap between inattentional blindness and attentional bias. Consciousness and

Muthén, L. K., & Muthén, B. O. (2017). Mplus User's Guide (8th ed.). Muthén & Muthén.

Performance, 10(1), 137–142. https://doi.org/10.1123/ijspp.2014-0048

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what is being reported? Criticisms and recommendations. Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Savelsbergh, G. J. P., Williams, A. M., Van der Kamp, J., & Ward, P. (2002). Visual search, anticipation and experience in football goalkeepers. Journal of Sports Sciences, 20(3), 279–287. https://doi.org/10.1080/026404102317284826

SEM. Journal of Strength and Fitness , 19(1), 231–240. https://doi.org/10.1519/15184.1

Streiner, D. L., Norman, G. R. & Cairney, J. (2015). Health measurement scales: a practical guide to their development and application (5th ed.). Oxford University Press.

Vaeyens, R., Lenoir, M., Williams, A. M., Mazyn, L., & Philippaerts, R. M. (2007). The effects of task-related constraints on visual search behaviour and decision-making ability in young footballers. Journal of Sport and Exercise Psychology, 29(2), 147–169. https://doi.org/10.1123/jsep.29.2.147

Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict success among footballers . PLOS ONE, 7(4), Article e34731. https://doi.org/10.1371/journal.pone.0034731

Ward, P., & Williams, A. M. (2003). Development of perceptual and cognitive skills in football: the multidimensional nature of elite performance. Journal of Sport and Exercise Psychology, 25(1), 93–111. https://doi.org/10.1123/jsep.25.1.93

Weir, J. P. (2005). Quantification of test-retest reliability using the intraclass correlation coefficient and the

Williams, A. M. & Reilly, T. (2000). Identification and development of talent in football. Journal of Sport Sciences, 18(9), 657–667. https://doi.org/10.1080/02640410050120041

Williams, A. M., & Burwitz, L. (1993). The use of anticipatory cues in football. In T. Reilly, J. Clarys & A. Stibbe (eds.), Science and football II (pp. 239–244). E & FN Spon.

Williams, A. M., & Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: some considerations on the application of the expert performance approach. Human Movement Science, 24(3), 283–307. https://doi.org/10.1016/j.humov.2005.06.002

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual-cognitive expertise in sport and its acquisition: implications for applied cognitive psychology. Applied Cognitive Psychology, 25(3), 432–442. https://doi.org/10.1002/acp.1710

Williams, A. M., Ward, P., Knowles, J. M. and Smeeton, N. J. (2004). Anticipatory ability in a real-world task: measurement, training and transfer in tennis. Journal of Experimental Psychology: Applied, 8(4), 259–270. https://doi.org/10.1037/1076-898X.8.4.259

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Young, W. B., Dawson, B., & Henry, G. J. (2015). Agility and speed of change of direction are independent abilities: implications for assessment and training. International Journal of Sports Physiology and