OPEN ACCESS
Turkish Journal of Kinesiology

ORIGINAL RESEARCH

Within-session test specific calibration in the T-agility test

Mürşit Aksoy[1], Davut Seyitoğulları[2], Yusuf Dilbilir[3]

[1] Department of Coaching Education, Faculty of Sports Sciences, Hakkari University, Hakkari, Türkiye. [2] Department of Physical Education and Sports, Institute of Graduate Education, Hakkari University, Hakkari, Türkiye. [3] Department of Medical Services and Techniques, Vocational School of Health Services, Hakkari University, Hakkari, Türkiye.

Received:
April 28, 2026

Accepted:
July 12, 2026

Published:
August 29, 2026

Keywords:
Familiarization, performance stabilization, practice effect, systematic bias, test specific calibration, T-agility test.

Abstract

This study aims to determine the minimum familiarization threshold (test-specific calibration) required to measure true performance capacity in the T-agility test and to mathematically demonstrate the systematic bias during this process. The high performance variability observed in the initial trials of agility tests stems from a cognitive and neuromotor calibration process where the athlete attempts to adapt to the test boundaries and rules, rather than from physical inadequacy. Thirty-four youth athletes with team-sport backgrounds (age: 16.12 ± 0.81 years) who were entirely naive to the standardized T-test protocol were included in the study. To optimally minimize the effects of fatigue, the athletes performed 9 consecutive successful T-test trials with 4-minute full recovery periods between each repetition. The obtained data were analyzed using Repeated Measures ANOVA, Intraclass Correlation Coefficient (ICC), and Bland-Altman analysis. The results indicated that the athletes' performance times improved significantly from the first trial (T1) to the sixth trial (T6). However, no statistically significant difference was detected among the T6, T7, T8, and T9 measurements, indicating that performance stabilized during this 4-trial plateau phase. A substantial systematic bias of 2.89 seconds was found between the participants' initial trial (T1) and their achieved peak score (Tpeak). Consequently, to allow athletes to reflect their true agility levels with optimally minimized fatigue and the novice reaction, we strongly recommend completing a calibration process of at least 5 repetitions at the beginning of the test. Based on these findings, the '5+3 Protocol'—which involves regarding the first 5 trials as familiarization to be excluded, and recording the peak score (best of the subsequent 3 measurement trials) to accurately reflect the athlete's true capacity is highly recommended for coaches and testing specialists to maximize measurement reliability.

Introduction

Agility is considered one of the most critical determinants of performance in complex sports disciplines. This skill requires a flawless neuromotor synchronization of acceleration, deceleration, and change of direction mechanics (Munro and Herrington, 2011). Fundamentally, agility is defined as the instantaneous reorganization of body posture and movement direction by generating a rapid response to a stimulus (Sheppard and Young, 2006). The constantly changing dynamics of the game in team and individual sports necessitate athletes to adapt their movements based on this foundation of agility (Morral-Yepes et al., 2022). Therefore, measuring this skill on the field with accurate, standardized, and reliable test protocols is of as much importance as developing agility through training.

The complex nature of agility is structurally divided into two primary components in the literature: Change of Direction (COD) speed and reactive agility. COD refers to the athlete's ability to change direction on a pre-planned route without the need for an external stimulus. This skill, relying entirely on physical motor execution, is generally classified as a 'closed skill' (Hülsdünker et al., 2026; Ye et al., 2025; Altmann et al., 2019). In contrast, reactive agility is a rapid and coordinated whole-body movement performed in

✉ M. Aksoy, e-mail: mursitaksoy@hakkari.edu.tr
Turk J Kinesiol, 12(4), 349-360. doi: 10.31459/turkjkin.1939112
Copyright: © 2026 by the author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by-nc/4.0/deed.en).

response to an external stimulus, such as an opponent's movement or a teammate's signal (Hülsdünker et al., 2026; Ye et al., 2025). The low to moderate correlations between COD speed and reactive agility tests prove that these two skills are entirely independent physical characteristics (Ye et al., 2025). Consequently, while COD tests (e.g., T-test, Illinois test) are specifically utilized to evaluate the physical capacity of directional changes, reactive agility tests provide an assessment closer to actual game scenarios (Friebe et al., 2024; Young et al., 2021; Özbay et al., 2018).

Contemporary agility paradigms emphasize that effective performance is not solely mechanical; it requires a highly complex integration of motor execution, perceptual-cognitive processes, and rapid decision-making under pressure (Espinosa et al., 2026). Although reactive agility tests successfully reflect field dynamics, they inherently contain visual and auditory 'cognitive noise' strictly dependent on the external stimulus. Because the stimulus or environmental conditions fluctuate with each repetition in these tests, it becomes exceedingly difficult to pinpoint the exact source of performance variations across consecutive trials. To isolate true physical motor adaptations from these unpredictable cognitive variables, recent exploratory factor analyses reveal that the T-test functions as a distinct performance construct separate from standard COD tests, primarily due to its inclusion of unique, sport-specific elements such as strictly controlled lateral and backward movements (Espinosa et al., 2026).

The T-agility test, one of the gold standards among agility tests, is a complex motor task with strict biomechanical constraints. These include touching the base of the cones, not crossing the feet during lateral shuffles, and maintaining a backward running posture (Pauole et al., 2000). Regardless of how high the athlete's motor capacity is, these specific constraints create a significant reliability problem in the initial stages of measurement. An athlete unfamiliar with the protocol expends a high level of cognitive effort. Their central nervous system must intensely work to integrate these sequential and distinct motor patterns. Consequently, the high performance variability observed in initial trials does not stem from physical inadequacy. Instead, it stems from the intense need for executive functions to accomplish the task according to the rules (Stuhr, Hughes, and Stöckel, 2018).

Times obtained in the initial trials by participants who are new to the test protocol contain a "systematic bias" and do not reflect their true motor capacities (Atkinson and Nevill, 1998; Hopkins, 2000). This bias is not merely a matter of the athlete trying to physically warm up to the measurement mechanics. Rather, the actual process is a "test-specific calibration" phase, where the athlete attempts to integrate their own sprint tempo, braking distance, and turning angles into the boundaries of the test. Consequently, initial measurements taken without adequate familiarization exhibit high test-retest variation, distancing performance from the true capacity (Stevens and Dascombe, 2015). Therefore, breaking this excessive variability in performance through consecutive repetitions in the T-test is a statistical and neuromotor necessity to isolate the athlete's true change-of-direction capacity.

Sensorimotor theories explain the rapid performance improvement in complex change-of-direction tasks through the instantaneous updating of internal models in the brain (Kawato, 1999; Shadmehr et al., 2010). According to motor control paradigms, when an organism is exposed to a new or restrictive environmental task, it rapidly constructs an internal model to predict the dynamics of the movement (Shadmehr, 2022; Wolpert et al., 1995). Schmidt's (1975) classic Schema Theory also supports this process; by receiving feedback on performance in each trial, the athlete readjusts their current motor program specifically for the task at hand. This process, which is often glossed over in traditional sports science literature with superficial terms such as "familiarization" or "practice effect" (Moir et al., 2004), is inherently a highly complex sensorimotor adaptation mechanism.

Some studies (Duncan et al., 2005; Hertel et al., 2000; Robinson and Gribble, 2008) demonstrate that the number of repetitions required for performance to stabilize in such high-intensity neuromotor tasks is directly related to the cognitive load and rule complexity inherent in the test, rather than the athlete's physical reserves. In this context, our study goes beyond the classic concept of familiarization and conceptualizes this rapid neural adaptation and error elimination process as "test-specific calibration." Considering the initial point where performance stabilizes as an absolute and insurmountable biological limit (asymptote) can be misleading from the perspective of motor control theory, as the plateau can be exceeded through methodological changes (Gray, 2017). Instead, it is a much more methodologically accurate approach to consider this state of equilibrium reached as a test-specific calibration or performance stabilization. Once the athlete reaches this plateau point, unnecessary steps

decrease, braking mechanics are optimized, and movement economy is maximized.

The most prominent methodological gap in the existing research specific to the T-Agility Test is that it has not been tied to a clear standard as to exactly where this 'test-specific calibration' phase ends and from which repetition performance stabilization begins. For example, although Pauole et al. (2000) reported high intraday reliability (R = 0.98) over three successful repetitions of the test, the study did not specify how many 'practice repetitions' participants were required to perform prior to the actual measurement. Similarly, Munro and Herrington (2011) suggested that a single trial application might be sufficient. However, their methodology presents serious limitations. The research was conducted over only four repetitions, leaving it ambiguous how scores would behave with additional trials. Most critically, they utilized a very short recovery period of just 1 minute between repetitions. To address this and ensure that physiological fatigue does not mask rapid neural adaptation, our proposed protocol implements a strictly controlled 4-minute recovery timeframe. Furthermore, the failure of past studies to explicitly report the participants' prior experiences (familiarization levels) with the test restricts the generalizability of the results.

Although it is generally assumed in the literature that 1-3 trials are sufficient, it is imperative to determine where the scores stabilize in individuals performing the test for the first time. Even though exercise durations and energy systems differ, the necessity of multiple trials for data stabilization is considered a methodological practice. This requirement is clearly observed in other anaerobic tests requiring maximal effort, such as the Wingate test (Barfield et al., 2002).

In alignment with these broader methodological principles, the intra-session reliability of agility tests is highly dependent on standardized protocols, consistent testing conditions, and the athletes' familiarity with specific movement constraints (Evrengul, 2023). Because noticeable initial learning effects frequently occur in such protocols, enhancing the familiarization of both athletes and testers is argued to be a critical requirement for maximizing measurement reliability (Ye et al., 2025). Without this necessary familiarization, high initial variability prevents the accurate assessment of the athlete's true physical capabilities, which inherently reinforces the need for a stable calibration threshold (Evrengul, 2023).

If this stabilization phase cannot be identified through objective analysis, the obtained measurements will not reflect the athlete's true agility. On the contrary, the recorded scores will measure the momentary "novice reaction" and cognitive confusion exhibited toward the test rules. The validity and reliability of the data to be obtained from agility tests depend on the accurate determination of the repetition at which the athlete reaches this plateau.

Based on this fundamental problem, the aim of the current study is to examine the test-specific calibration process in club-level competitive youth athletes aged 15–17 years from various team sports (basketball, volleyball, and soccer) who have a training background but are specifically unfamiliar with the T-test protocol (familiarization-naive). Accordingly, it is aimed to determine the optimal number of repetitions at which performance reaches a plateau and to establish the most reliable measurement protocol that will reflect the athlete's true baseline capacity.

Within the framework of existing motor control theories and the current performance literature, the main hypothesis of our research is as follows: Participants will exhibit a high systematic bias due to coordination deficiencies in adjusting cone contact points, lowering their center of gravity, and optimizing deceleration strategies during the initial phase of the test. Contrary to traditional protocols suggesting 1-3 trials, it is predicted that athletes will require more than 3 repetitions for the 'test-specific calibration' process to reach a true performance stabilization phase. Ultimately, we hypothesize that measurements taken during this final plateau phase will exhibit high internal consistency. Free from the effects of novelty and cognitive load, these stabilized scores will reflect the athlete's true neuromotor capacity.

Methods

This study employed a quantitative, single-group repeated measures design. The objective of this design was to systematically examine the within-session, test-specific calibration process and performance stabilization in the T-agility test through nine consecutive trials administered under standardized conditions.

Participants

Thirty-four male athletes, who actively participate in competitions at national and regional levels in various disciplines (football, basketball, and volleyball) within a

regional sports directorate and possess at least two years of regular training experience, were included in the study through the convenience sampling method. Although this sampling approach presents a minor limitation regarding broad generalizability, it is well-suited for controlled fieldwork. The diversity of branches was kept heterogeneous to reflect the multifaceted nature of the T-agility test and its universal applicability across different sports disciplines. The demographic and physical characteristics of the participants are as follows: Age: 16.12±0.81 years; Height: 173.29±6.73 cm; Body Weight: 61.65±11.71 kg; and Body Mass Index (BMI): 20.44±2.82 kg/m² (Table 1). The inclusion criteria for the study were the absence of any known lower extremity injuries and being previously unexposed to the T-agility test protocol within the scope of a standardized measurement (familiarization-naive). All participants and the parents of minors were informed about the study procedures, and written informed consent was obtained. The study was approved by the Hakkari University Scientific Research and Publication Ethics Committee (2024/229).

Procedure

The measurements were conducted on a polyurethane surface in a closed sports hall, isolated from external factors (wind, noise, spectators). All participants refrained from strenuous physical activity for 24 hours prior to the testing session and arrived for the measurements in a rested state. Prior to the test, a standardized 15-minute sub-maximal warm-up protocol was administered, consisting of straight-line running, dynamic stretching, and lateral shuffle movements consistent with the mechanics of the T-test. The T-agility test was then administered in strict adherence to the original protocol described by Semenick (1990). The athletes completed the course—which involves a 9.14-meter straight sprint, followed by 4.57-meter lateral shuffles to the right and left, and a 9.14-meter backward run to the starting point—in the shortest possible time. During the measurement process, several strict control mechanisms were observed. Specifically, a 4-minute rest period was administered between trials in accordance with the literature to ensure the complete recovery of the phosphagen (ATP-PC) system and to eliminate fatigue-induced performance decrements as a confounding variable (Bogdanis et al., 1995). Analyses were conducted based on nine consecutive successful trials that contained no errors, such as knocking over cones, crossing of feet, or failing to touch the targets. Furthermore, to prevent a competitive environment, blinding was implemented by withholding the photocell results from the athletes, which were recorded solely by the evaluators. Finally, in the event of failed repetitions, the trial was immediately halted. The athlete was required to repeat the attempt only after completing a full 4-minute recovery period to prevent fatigue accumulation. Athletes were only provided with performance feedback regarding rule violations, and any motivational or tactical guidance was strictly avoided.

Data Collection Tools

The body weight measurements of the participants were recorded using a Tanita DC 360 scale. Height measurements were obtained using a wall-mounted tape measure with a precision of 0.1 cm and a spirit level apparatus. The T-agility test performance times were recorded using a wireless photocell timing system with a precision of 0.001 seconds (Gama Technology, Turkiye). The photocell gates were positioned at the starting line at a height of 1 meter from the ground, corresponding to the athletes' trunk height.

Data Analyses

Data analysis was performed using the SPSS version 26.0 statistical software package. The normal distribution of the data was assessed using the Shapiro-Wilk test. A Repeated Measures ANOVA was utilized to determine the performance change across the nine consecutive trials and to identify the stabilization phase. A Bonferroni post-hoc analysis was applied to determine the source of any significant differences. The reliability of the measurements during the phase where performance stabilized (plateau) was tested using the Intraclass Correlation Coefficient (ICC) and the Coefficient of Variation (CV). dditionally, the relationship between the best score (Tpeak) and the plateau average was examined using Pearson Correlation. Furthermore, Multiple Linear Regression was utilized to determine whether age and BMI acted as confounding variables within this homogenous group. The level of statistical significance was set at $p < 0.05$.

The sample size and statistical power of the study were evaluated via post-hoc analysis using G*Power software (v3.1.9.7). Given a medium effect size ($f = 0.25$), a significance level of 0.05 (α), and the parameters of 9 repeated measurements; the statistical power ($1-\beta$) of the 34-person sample was calculated to be 0.995. This value is well above the minimum power level of 0.80 required to detect differences between trials, demonstrating that the obtained findings possess high reliability.

Results

The descriptive statistics of the 34 athletes participating in the study are presented in Table 1. The mean age of the participants was 16.12 ± 0.81 years (Min: 15, Max: 17). Regarding physical characteristics, the participants had a mean height of 173.29 ± 6.73 cm and a mean body weight of 61.65 ± 11.71 kg. BMI, calculated from the obtained height and weight values, was determined to be 20.44 ± 2.82 kg/m².

A Repeated Measures ANOVA was applied to compare the performance times of the participants across 9 consecutive T-agility test trials (T1-T9) (Table 2). Mauchly's test indicated that the assumption of sphericity had been violated ($p < .001$); therefore, the degrees of freedom were corrected using the Greenhouse-Geisser method. The analysis revealed a highly statistically significant main effect across the test trials ($F = 95.14$, $p < .001$, $\eta p^2 = .742$). Bonferroni-corrected pairwise comparisons demonstrated that performance times decreased (i.e., improved) significantly from T1 to T6 ($p < .05$). However, no statistically significant difference was detected among the T6, T7, T8, and T9 measurements ($p > .05$). The Familiarization curve is illustrated in Figure 1.

Table 1
Descriptive characteristics of the participants (n=34).

Variables	Mean	Std. Dev.	Minimum	Maximum
Age (years)	16.12	0.81	15.00	17.00
Height (cm)	173.29	6.73	161.00	190.00
Weight (kg)	61.65	11.71	45.00	100.00
BMI (kg/m^2)	20.44	2.82	16.89	29.39

As shown in Table 3, an Intraclass Correlation Coefficient (ICC) analysis was performed to evaluate the internal consistency among the 6th, 7th, 8th, and 9th repetitions (T6, T7, T8, T9), where the calibration phase was completed. The average measures ICC value, calculated using a two-way mixed-effects model and the definition of absolute agreement, was found to be .870 (95% CI [.781, .929], $p < .001$). Furthermore, the relationship between the average of these four repetitions (TmPlat) and the peak score (Tpeak) recorded during the session was examined using Pearson correlation analysis. he analysis revealed a very strong positive and significant correlation between the TmPlat and Tpeak variables ($r = .927$, $p < .001$). This strong linear relationship is visually illustrated in the scatter plot (Figure 2).

Table 2
Repeated measures ANOVA results for T-agility test Trials.

Trial	Mean	SD	F	p	η_p^2
T1	14.03^a	1.16			
T2	13.35^b	0.93			
T3	12.91^{bc}	0.73			
T4	12.45^{cd}	0.71			
T5	12.22^d	0.65	95.14	< .001	.742
T6	11.66^e	0.58			
T7	11.42^e	0.58			
T8	11.54^e	0.65			
T9	11.53^e	0.60			

SD: Standard Deviation; η_p^2: Partial Eta Squared.
Means sharing the same superscript letter (a, b, c, d, e) are not significantly different from each other ($p > .05$) based on Bonferroni adjusted pairwise comparisons. Degrees of freedom for the F-ratio were adjusted using the Greenhouse-Geisser estimate ($\epsilon = .580$) as the assumption of sphericity was violated.

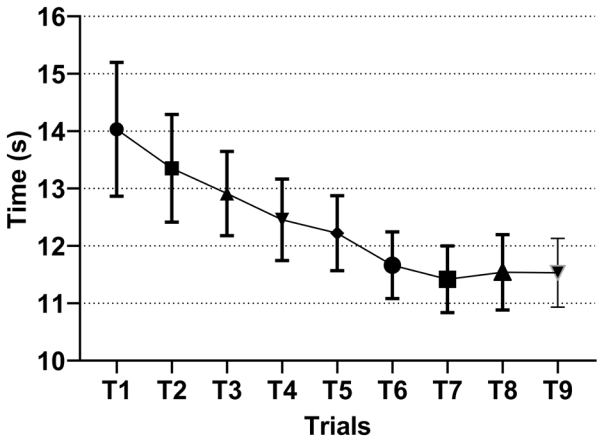


Figure 1. Familiarization curve T-agility test trials.

Table 3
Reliability and correlation analysis of the stabilization phase (Trials 6-9).

Measure	Statistic	Value	95% CI	p
T6, T7, T8, and T9 Consistency	ICC	.870	[.781, .929]	< .001
Tpeak vs. TmPlat	Pearson r	.927	-	< .001

CI: Confidence Interval; Tpeak: Min. of T1 to T9; TmPlat: Mean of the plateau phase (T6, T7, T8, and T9).

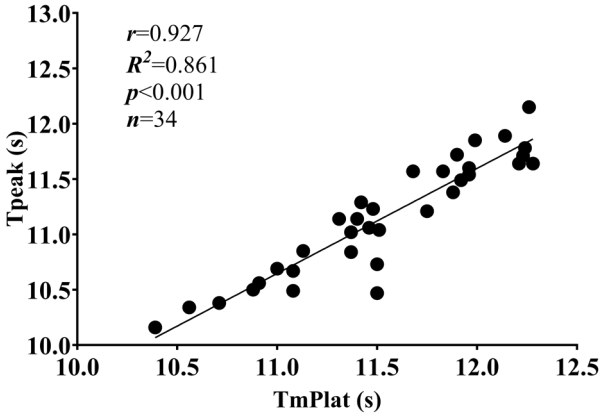


Figure 2. Familiarization curve T-agility test trials.

A multiple linear regression analysis was conducted to predict the amount of score improvement (T1 - Tpeak) based on the Age and BMI values of the participants (Table 4). The formulated regression model was found to be statistically non-significant ($F(2, 31) = 0.841$, $p = 0.441$), with an explanatory variance of $R^2 = 0.051$ and an Adjusted R^2 of -0.010. The results of the analysis indicated that neither Age ($\beta = -0.112$, $p = 0.530$) nor BMI ($\beta = 0.182$, $p = 0.311$) emerged as significant predictors of the amount of test-specific calibration**.** The negative Adjusted R^2 indicates that these variables do not provide additional predictive value for this model. This outcome can be attributed to the demographic homogeneity of the sample. The participants are adolescent athletes with very similar age and physical characteristics. Due to this restricted variance, it is statistically challenging for Age and BMI to act as strong predictors within this specific cohort.

A Bland-Altman analysis was conducted to evaluate the systematic bias and 95% limits of agreement (LoA) between consecutive T-agility test trials (Table 5). According to the analysis results, the mean systematic bias between the participants' initial trial (T1) and their achieved peak performance (Tpeak) was calculated as 2.89 s (SD = 1.03), and the 95% limits of agreement were found to range from 0.869 to 4.91 s. When consecutive trials were examined, the systematic bias between T1 and T2 at the onset of the calibration phase was observed to be 0.681 s (95% LoA: -1.41 to 2.77). The bias between the fifth and sixth trials (T5-T6), immediately preceding the plateau, decreased to 0.559 s (SD = 0.603; 95% LoA: -0.624 to 1.74). Throughout the plateau (stabilization) phase, this bias continued to progressively attenuate; it was recorded as 0.245 s (SD = 0.537; 95% LoA: -0.808 to 1.30) between the sixth and seventh trials (T6-T7), and as -0.122 s (SD = 0.478; 95% LoA: -1.06 to 0.816) between the seventh and eighth trials (T7-T8). Finally, the systematic bias between the eighth and ninth trials (T8-T9), representing the culmination of performance stabilization, was practically neutralized, dropping to 0.008 s (SD = 0.430), and the limits of agreement narrowed to a range of -0.836 to 0.851 s.

Table 4
Multiple linear regression analysis predicting score improvement (calibration amount) from age and BMI.

	B	SE	β	t	p	VIF
(Constant)	3.841	4.045		0.950	.350	-
Age	-0.143	0.226	-.112	-0.636	.530	1.020
BMI	0.067	0.065	.182	1.030	.311	1.020

$R^2 = .051$; Adjusted $R^2 = -.010$; $F(2, 31) = 0.841$; $p = .441$; Dependent Variable: Score Improvement (T1 - Tpeak).
BMI: Body Mass Index; VIF: Variance Inflation Factor.

Table 5
Bland-Altman analysis results between consecutive T-agility test trials.

Comparison	Bias (s)	SD of Bias	95% LoA
T1 - T2	0.681	1.07	-1.41 to 2.77
T5 - T6	0.559	0.603	-0.624 to 1.74
T6 - T7	0.245	0.537	-0.808 to 1.30
T7 - T8	-0.122	0.478	-1.06 to 0.816
T8 - T9	0.008	0.430	-0.836 to 0.851
T1 - Tpeak	2.89	1.03	0.869 to 4.91

SD: Standard Deviation; LoA: Limits of Agreement; Tpeak: The best score achieved during the 9-trial session.
All units are in seconds (s). T8-T9 represents the extreme end of the stabilization plateau phase.

Discussion

The primary objective of this study is to determine the minimum familiarization threshold required for the stabilization of true performance in the T-agility test and to mathematically demonstrate the magnitude of methodological bias during this process. The obtained findings demonstrate a systematic bias of 2.89 seconds between the participants' initial trials (T1) and their achieved peak scores (Tpeak). This difference challenges the validity of the "single or low-repetition measurement" approaches widely accepted in the agility literature. Bland-Altman analyses revealed that, contrary to the 2-3 trial protocols deemed sufficient in traditional literature, significant improvements continued during the initial five repetitions. Consequently, the true plateau (performance stabilization) was achieved starting from the 6th trial. The statistical robustness of these findings is supported by a high power level of 99.5% (Power = 0.995). This high statistical power confirms that the identified "test-specific neuromotor calibration" process is a methodological reality free from random errors. Therefore, it is concluded that the short familiarization protocols of 2-3 trials frequently recommended for the T-agility test in the literature (Pauole et al., 2000; Munro and Herrington, 2011) are insufficient in light of the post-hoc analysis results obtained in this study. Consequently, relying on shortened test protocols in the absence of proper structural validation and performance stabilization can severely compromise the interpretability of results, failing to reflect the athlete's actual cognitive-motor capacities (Espinosa et al., 2026).

A key finding of the study is the systematic bias of 0.559 seconds observed between the T5 and T6 trials. This difference, exceeding half a second in an agility test lasting approximately 10-12 seconds, indicates that athletes continue to refine their neuromotor strategies (braking, changing direction) through the 5th trial. Consequently, they have not yet reached the stabilization threshold during these initial attempts. A variance of this magnitude exceeds typical measurement error margins and highlights a potential methodological validity issue in shorter protocols. This calibration phase, where the athlete instantaneously updates their internal model according to the rules of the test, concludes by the 6th repetition. To obtain reliable data, administering a 5-trial familiarization process prior to 3 actual measurement trials (conceptualized herein as the '5+3 Protocol') is recommended. This aligns with recent evidence emphasizing that because noticeable initial learning effects frequently occur in test protocols, enhancing the familiarization of both athletes and test administrators is a critical requirement to maximize measurement reliability (Friebe et al., 2023). Ultimately, achieving adequate absolute reliability and ensuring sufficient sensitivity to performance changes dictate the implementation of a separate, comprehensive familiarization phase regarding the test layout (Friebe et al., 2023).

Another key finding of the study is that, according to the regression analysis results, the participants' age and Body Mass Index (BMI) values had no predictive effect on the 'amount of improvement' (calibration level) (R^2 = 0.051, p = 0.441). Despite a wide range in the participants' body weight (45 kg to 100 kg) and BMI (16.89 to 29.39 kg/m²) values, the non-significant model indicates that the process of adapting to the test course is a neuromotor adaptation independent of inter-individual morphological differences. Although the systematic bias of 2.89 seconds detected between T1 and Tpeak is specific to the athletically inclined youth group of 34 individuals in this research, the observed decrease in error as the number of trials increases reflects a fundamental pattern of motor performance stabilization.

This neuromotor adaptation is also supported by the high internal consistency coefficient (ICC = 0.870; 95% CI [0.781, 0.929]) detected among the 6th, 7th, 8th, and 9th repetitions. During these trials, the calibration phase was already completed. This consistency documents that the athletes fully adapted to the test mechanics and the cognitive 'novice' effect diminished. On the other hand, in a test design with multiple repetitions (8-9 trials), the suspicion may arise that this pause in performance (plateau) could stem from 'fatigue accumulation.' However, the 4-minute active rest period applied in the research is thought to have provided a sufficient interval for the resynthesis of phosphagen (ATP-PC) stores. This protocol effectively mitigated cumulative fatigue. Indeed, if the performance plateau had stemmed from fatigue rather than the completion of neuromotor calibration, the consistency between measurements would have markedly declined. Yet, the very strong positive correlation (r = 0.927, p < 0.001) between the average of the plateau phase (TmPlat) and the recorded peak score (Tpeak) indicates otherwise. This correlation demonstrates that the achieved scores are not a result of fatigue or random variance. On the contrary, this plateau is a valid parameter that provides

a stable and accurate representation of the athlete's true baseline capacity at that moment.

When studies on the T-Agility Test are examined, it is evident that researchers have largely focused on the 'reliability' of the test. However, they have not presented a standardized framework regarding exactly after how many repetitions performance stabilizes (Munro and Herrington, 2011; Pauole et al., 2000; Sassi et al., 2009). Furthermore, methodological inconsistencies exist in the literature. The conversion of the 5 yd–10 yd measurements in Semenick's (1990) original protocol to the metric system as 5 m–10 m in some studies has weakened comparability across studies (Miller et al., 2006). In addition, variations in data processing strategies are also notable. For instance, some studies included the best two trials, some the average of three trials, and others an unspecified number of familiarization trials in their analyses (Ahmed et al., 2022; Armstrong and Greig, 2018; Chaabene et al., 2018; Gisladottir et al., 2024; Lindblom et al., 2020; Munro and Herrington, 2011; Pauole et al., 2000; Rey et al., 2024). Therefore, implementing standardized measurement designs based on rigorous protocols is an essential requirement to ensure reliable comparability among studies.

A prominent example of these limitations in the literature is seen in the study by Munro and Herrington (2011). In that research, it was reported that performance stabilization in the T-Agility Test occurred at the third repetition for males and the fourth repetition for females. However, the researchers terminated the testing protocol at this point and could not observe how performance would develop during subsequent trials (e.g., the continued improvement through the 5th repetition observed in our study). More importantly, it is thought that the premature stabilization and performance fluctuations observed in Munro's research stemmed from the short 1-minute rest period administered between trials. A 1-minute recovery following intense exercise allows for the replenishment of only 50–70% of creatine phosphate (CP) stores. In our research, however, providing at least 4 minutes of active rest between trials is considered to have facilitated the replenishment of more than 90% of CP stores (Tahiroğlu and Shahidi, 2025). Through this physiological framework, the consecutive performance improvement observed in our study was differentiated from cumulative fatigue. This approach effectively isolated the test-specific neuromotor adaptation (calibration) process.

On the other hand, Oliveira et al. (2024) suggested that the familiarization phase could be neglected in simple tests that do not involve complex motor skills. However, this approach is only valid for linear tests with low cognitive load. It has been stated that in complex skill tests containing multidirectional movements, changes of direction, and strict postural constraints, such as the T-Agility Test, data obtained without considering the practice effect will not reflect true capacity (Bekris et al., 2018; Benvenuti et al., 2010).

Due to the limited direct data on performance stabilization in the T-Agility Test literature, studies examining calibration processes in tests requiring complex motor skills and coordination offer an important basis for comparison. In this context, the studies by Hertel et al. (2000) on the Y Balance Test stand out. In their studies examining intra-rater and inter-rater reliability, the researchers reported that participants reached their best performance values only between the 7th and 9th repetitions. Deepening this finding further, Robinson et al. (2008) stated that the stabilization process was delayed due to the increase in the "biomechanical degrees of freedom" that needed to be coordinated in conditions where the arms were free; whereas in simpler situations where the arms were kept fixed, stabilization began around the 4th repetition.

This situation clearly shows that as the complexity of the motor tasks within the test protocol increases, the test-specific calibration process also lengthens. Similarly, Duncan et al. (2005) examined the practice effect in backward medicine ball throws and reported that optimal stabilization could only be achieved after 5-6 repetitions. It was observed that typical errors in performance decreased as the number of repetitions increased, and it was emphasized that the stabilization phase could be demonstrated more soundly with a sufficient number of repetitions. Although the T-Agility test and balance or medicine ball throws require completely different motor skills, the common ground of these tests is their high biomechanical complexity. This explicitly demonstrates that as the complexity of motor tasks increases, the test-specific calibration process is correspondingly prolonged. Considering the nature of the T-Agility test, which involves multiple changes of direction, sudden braking, and strict postural constraints (high biomechanical degrees of freedom), it is proven that the 5-repetition calibration necessity identified in our study parallels the requirements of these complex motor tests in the literature.

Due to the nature of this neuromotor adaptation process, the profile of the participants is also of great importance. It is known that individuals with high bodily-kinesthetic intelligence potential adapt to new motor skills more quickly and can successfully adapt their past motor experiences to new tasks (Davis et al., 2011; Di Domenico et al., 2019). However, in order to clearly reveal the actual neuromechanical dynamics and the performance stabilization process, it is necessary to examine specifically naive individuals (Guadagnoli and Lee, 2004). For this reason, participants coming from team sports (possessing high motor potential) but completely unfamiliar with the T-Agility Test protocol were preferred in our research. This methodological preference allowed us to obtain a clear and isolated version of the practice effect during the detection of the calibration phase. Although it is predicted that this test-specific calibration time may be shorter in elite-level groups, the minimum 5-trial protocol we determined serves as a strong baseline rather than an absolute law, offering a highly valid framework for similar demographics. Furthermore, the interpretation of these stabilization thresholds must explicitly account for the specific demographic characteristics of the sample. Fundamental factors such as biological maturation, neuromotor development, and age differences lead to pronounced variations in agility scores, particularly among youth athletes (Ye et al., 2025). Additionally, sex significantly impacts intra-session reliability metrics, as female participants typically exhibit lower measurement error and better reliability values compared to males (Doma et al., 2023). Therefore, the findings of this study should be evaluated within the context of male youth team-sport athletes, acknowledging that test-specific calibration parameters may differ across other athletic populations.

In the literature, it has been shown that familiarization trials substantially reduce variation in test-retest measurements. If the performance increase in the initial trials stemmed solely from a physical warm-up effect, measurement variation would be expected to rise again after extended rest days. However, even with recovery periods of up to seven days between tests, the familiarization factor prevents these measurement biases (Stevens and Dascombe, 2015). This indicates that the adaptation achieved through repetitions is not purely physical arousal. On the contrary, this improvement is a structural adaptation process primarily based on procedural, cognitive, and test-specific calibration mechanisms.

Although these findings have predominantly been obtained through different exercise protocols, the reduction of measurement variation along with familiarization is a fundamental concept of motor control theory. Nevertheless, the fact that performance stabilization specific to the T-Agility test in the current research was determined solely through intra-day repetitions is a methodological limitation. It is necessary to verify the complete retention of this temporary calibration dynamic, observed in other tests in the literature, in the T-Agility test. In this context, it is recommended that future studies construct inter-day or longer-interval test-retest designs and examine the rate of decay of the reached plateaus over time.

Limitations

While this study contributes to the understanding of the test-specific calibration process, several limitations must be considered when interpreting the findings. First, sample size considerations (n = 34) and the relatively specific sample characteristics, consisting exclusively of an athletically inclined demographic of male athletes aged 15-17 years, must be acknowledged. Additionally, although the chronological ages of the participants were recorded, the potential effects of growth and biological maturation rates on neuromotor adaptation processes were not controlled. Furthermore, the absence of external validation measures (e.g., 3D kinematics or laboratory-based biomechanical assessments) limits the ability to cross-verify the identified plateau phase. This demographic structure and methodological scope restrict the generalizability of the findings to other age groups, competitive levels, sexes, or sedentary individuals. Similarly, because the participants came from different sports disciplines (volleyball, basketball, football), the specific effects of branch-specific motor habits on test adaptation were not isolated. However, it should be noted that the main focus of the research is not an 'inter-branch performance comparison' but rather the 'test-specific calibration process.' The fact that such a consistent plateau curve was obtained and a substantial performance difference was detected even in a heterogeneous sample group supports that the proposed calibration protocol serves as a strong baseline rather than an absolute law.

A second limitation is that the study focused solely on the T-Agility test. The potential learning effects during repeated trials and the number of repetitions required for performance stabilization may vary in tests with different biomechanical requirements, turning

angles, and rule sets, such as the Illinois, 5-0-5, or Pro-Agility tests. It is a critical step for future research to adapt this stabilization analysis to other popular field tests in order to resolve the standardization issues in the literature.

Ultimately, the fact that the plateau specific to the T-Agility test in the research was determined solely through intra-day repetitions is a significant time factor limitation. In the literature, it has been shown that the effect of the familiarization factor in reducing measurement variation is maintained even after long recovery periods; that is, this adaptation is a cognitive and procedural adjustment rather than mere physical warm-up (Stevens and Dascombe, 2015). To address this limitation and provide a clear roadmap for future kinesiology research, the following future research directions are proposed. Specifically, replication in larger and more diverse cohorts, as well as direct comparisons across sexes and performance levels, are strongly recommended. Moreover, it is crucial to conduct investigations of between-session reliability by constructing longer-interval (weekly) test-retest designs, and to evaluate the responsiveness of this test-specific calibration process to various training interventions.

Conclusion

This study strongly suggests that the marked performance improvement observed in the initial phase of the T-Agility test is a dynamic 'test-specific calibration' process wherein systematic bias is eliminated, rather than permanent motor learning. The obtained findings demonstrate that even in youth athletes with team-sport backgrounds, the nervous system requires a test-specific calibration phase to adapt to the specific constraints and course geometry of the test. The statistically determined 5th-repetition threshold is the critical point at which performance stabilizes and reaches a plateau. It has been demonstrated that data obtained prior to this threshold are far from reflecting the athlete's true capacity and lead to a methodological bias. Consequently, for the T-Agility test to be considered a valid parameter of physical performance, the completion of a minimum 5-repetition calibration phase is recommended.

Practical Applications

The obtained findings provide a practical and highly reliable measurement standard that accounts for time and energy economy for coaches, strength and conditioning professionals, and sports scientists working in the field. In this context, the '5+3 Protocol' is recommended to ensure that agility performance is measured while optimally minimizing fatigue and the novice reaction. Specifically, in T-Agility test administrations, the first 5 trials must be regarded as 'familiarization/calibration' repetitions and excluded from the final evaluation. Following these initial repetitions, 3 actual measurement trials should be conducted. Limiting this measurement phase to exactly 3 trials ensures maximum time and energy economy in field settings, preventing unnecessary fatigue while remaining statistically sufficient to capture the athlete's true peak capacity. The best score (peak score) obtained from these 3 trials should be recorded as the athlete's true agility level. As demonstrated by the strong correlation between the plateau average and peak scores in our findings, recording the peak score ensures both physiological validity and alignment with standard field-testing practices. To maximize measurement reliability during this protocol, a 4-minute full rest period must be provided between trials to optimally minimize the effect of fatigue. Additionally, the strategy of withholding trial results from the athlete (blinding) should be maintained as a standard practice to minimize motivational fluctuations and systematic bias. Furthermore, for future research, determining the calibration thresholds in other agility tests with different biomechanical constraints (e.g., Illinois, 5-0-5) and examining the rates of neuromotor adaptation across genders will help resolve the existing standardization issues in the literature.

Authors' Contribution

Study Design: MA, DS; Data Collection: MA, DS; Statistical Analysis: YD; Manuscript Preparation: MA, DS; Funds Collection: MA.

Ethical Approval

The research protocol was formally reviewed and approved by the Hakkari University Scientific Research and Publication Ethics Committee (Approval Date: December 6, 2024; Session Date: December 5, 2024; Session Number: 2024/229; Committee Decision No: 1). and it was carried out in accordance with the Code of Ethics of the World Medical Association also known as a declaration of Helsinki.

Funding

This work was supported by the Hakkari University Scientific Research Project Unit under Project Number: SB25BAP1.

Conflict of Interest

The authors hereby declare that there was no conflict of interest in conducting this research.

References

Ahmed, S., Saraswat, A., & Esht, V. (2022). Correlation of core stability with balance, agility and upper limb power in badminton players: A cross-sectional study. *Sport Sci Health*, 18(1), 165–169. doi: 10.1007/s11332-021-00789-w

Altmann, S., Ringhof, S., Neumann, R., Woll, A., & Rumpf, M. C. (2019). Validity and reliability of speed tests used in soccer: A systematic review. *PloS One, 14*(8), e0220982. doi: 10.1371/journal.pone.0220982

Armstrong, R., & Greig, M. (2018). The Functional Movement Screen and modified Star Excursion Balance Test as predictors of T-test agility performance in university rugby union and netball players. *Phys Ther Sport*, *31*, 15–21. doi: 10.1016/j.ptsp.2018.01.003

Atkinson, G., & Nevill, A. M. (1998). Statistical Methods For Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine. *Sports Med*, *26*(4), 217–238. doi: 10.2165/00007256-199826040-00002

Barfield, J.-P., Sells, P. D., Rowe, D. A., & Hannigan-Downs, K. (2002). Practice effect of the Wingate anaerobic test. *J Strength Cond Res*, *16*(3), 472-473.

Bekris, E., Gissis, I., & Kounalakis, S. (2018). The dribbling agility test as a potential tool for evaluating the dribbling skill in young soccer players. *Res Sports Med*, *26*(4), 425–435. doi: 10.1080/15438627.2018.1492395.

Benvenuti, C., Minganti, C., Condello, G., Capranica, L., & Tessitore, A. (2010). Agility assessment in female futsal and soccer players. *Medicina*, *46*(6), 415. doi: 10.3390/medicina46060058

Bogdanis, G. C., Nevill, M. E., Boobis, L. H., Lakomy, H. K., & Nevill, A. M. (1995). Recovery of power output and muscle metabolites following 30 s of maximal sprint cycling in man. *J Physiol*, *482 (Pt 2)*, 467–480. doi: 10.1113/jphysiol.1995.sp020533

Carvajal-Espinosa, R., Salazar-Rojas, W., Navarro-Martínez, C., & Pino-Ortega, J. (2026). Assessing validity of agility and change of direction tests in soccer players: A factor analysis. *Journal of Physical Education and Sport, 26*(4), 666-674. doi: 10.7752/jpes.2026.04067

Chaabene, H., Negra, Y., Capranica, L., Bouguezzi, R., Hachana, Y., Rouahi, M. A., & Mkaouer, B. (2018). Validity and reliability of a new test of planned agility in elite taekwondo athletes. *J Strength Cond Res*, *32*(9), 2542–2547. doi: 10.1519/JSC.0000000000002325

Davis, K., Christodoulou, J., Seider, S., & Gardner, H. E. (2011). *The theory of multiple intelligences* [SSRN Scholarly Paper]. Social Science Research Network. https://papers.ssrn.com/abstract=2982593

Di Domenico, F., Fattore, S., Pignato, S., & D'Isanto, T. (2019). Relationship between motor learning and reaction capacity in motor task. *Journal of Human Sport and Exercise*, 14(4proc), S1030-S1037. doi: 10.14198/jhse.2019.14.Proc4.65

Doma, K., Connor, J. D., Nakamura, F. Y., & Leicht, A. S. (2023). Intra-session reliability of sprint performance on a non-motorised treadmill for healthy active males and females. *J Hum Kinet, 88*, 163–171. doi: 10.5114/jhk/163180

Duncan, M. J., Al-Nakeeb, Y., & Nevill, A. M. (2005). Influence of familiarization on a backward, overhead medicine ball explosive power test. *Res Sports Med*, *13*(4), 345–352. doi: 10.1080/15438620500359950

Evrengul, M. C. (2023). *Developing new reactive agility tests in basketball* (Doctoral dissertation, Lietuvos Sporto Universitetas). https://vb.lsu.lt/object/elaba:167150990/

Friebe, D., Hülsdünker, T., Giesche, F., Banzer, W., Pfab, F., Haser, C., & Vogt, L. (2023). Reliability and usefulness of the SKILLCOURT as a computerized agility and motor–cognitive testing tool. Med Sci Sports Exerc, 55(7), 1265-1273. doi: 10.1249/MSS.0000000000003153

Friebe, D., Sieland, J., Both, H., Giesche, F., Haser, C., Hülsdünker, T., ... & Banzer, W. (2024). Validity of a motor–cognitive dual-task agility test in elite youth football players. *Eur J Sport Sci, 24*(8), 1056-1066. doi: 10.1002/ejsc.12153

Gisladottir, T., Petrović, M., Sinković, F., & Novak, D. (2024). The relationship between agility, linear sprinting, and vertical jumping performance in U-14 and professional senior team sports players. *Front Sports Act Living*, *6*, 1385721. doi: 10.3389/fspor.2024.1385721

Gray, W. D. (2017). Plateaus and asymptotes: spurious and real limits in human performance. *Curr Dir Psychol Sci*, *26*(1), 59–67. doi: 10.1177/0963721416672904

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav*, *36*(2), 212–224. doi: 10.3200/JMBR.36.2.212-224

Hertel, J., Miller, S. J., & Denegar, C. R. (2000). Intratester and intertester reliability during the Star Excursion Balance Tests. *J Sport Rehabil*, *9*(2), 104–116. doi: 10.1123/jsr.9.2.104

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Med*, *30*(1), 1–15. doi: 10.2165/00007256-200030010-00001

Hülsdünker, T., Bahdur, K., Flammang, L., Naclerio, F., Sondermann, Y., Mierau, A., & Karsten, B. (2026). Assessing maximum oxygen uptake through a motor-cognitive reactive agility test in team ball sports athletes. *Front Sports Act Living*, *8, 1749132.* doi: 10.3389/fspor.2026.1749132.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr Opin Neurobiol*, *9*(6), 718–727. doi: 10.1016/S0959-4388(99)00028-8

Lindblom, H., Waldén, M., & Hägglund, M. (2020). Performance effects with injury prevention exercise programmes in male youth football players: a randomised trial comparing two interventions. *Sports Med Open*, *6*(1), 56. doi: 10.1186/s40798-020-00282-7

Miller, M. G., Herniman, J. J., Ricard, M. D., Cheatham, C. C., & Michael, T. J. (2006). The effects of a 6-week plyometric training program on agility. *J Sports Sci Med*, *5*(3), 459–465.

Moir, G., Button, C., Glaister, M., & Stone, M. H. (2004). Influence of familiarization on the reliability of vertical jump and acceleration sprinting performance in physically active men. *J Strength Cond Res*, *18*(2), 276–280. doi: 10.1519/R-13093.1

Morral-Yepes, M., Moras, G., Bishop, C., & Gonzalo-Skok, O. (2022). Assessing the reliability and validity of agility testing in team sports: a systematic review. *J Strength Cond Res*, *36*(7), 2035. doi: 10.1519/JSC.0000000000003753

Munro, A. G., & Herrington, L. C. (2011). Between-session reliability of four hop tests and the agility t-test. *J Strength*

Cond Res, *25*(5), 1470-1477. doi: 10.1519/JSC.0b013e3181d83335

Oliveira, D. C. X. de, Gonçalves, H. R., Denardi, R., Oliveira, V., & Deminice, R. (2024). The influence of familiarization on performance testing of young soccer players. *Motriz Revista de Educação Física*, *30*(1), e10210188. doi: 10.5016/s1980-6574e10210188

Özbay, S., Ulupınar, S., & Özkara, A. B. (2018). Agility performance in sports. *Journal of National Sport Sciences, 2*(2), 97-112. doi: 10.30769/usbd.463364

Pauole, K., Madole, K., Garhammer, J., Lacourse, M., & Rozenek, R. (2000). Reliability and validity of the t-test as a measure of agility, leg power, and leg speed in college-aged men and women. *J Strength Cond Res*, *14*(4), 443-450.

Rey, E., Carrera, S., Padrón-Cabo, A., & Costa, P. B. (2024). Effectiveness of short vs. long-distance sprint training on sprinting and agility performance in young soccer players. *Biol Sport*, *41*(1), 87–93. doi: 10.5114/biolsport.2024.127384

Robinson, R. H., & Gribble, P. A. (2008). Support for a reduction in the number of trials needed for the Star Excursion Balance Test. *Arch Phys Med Rehabil*, *89*(2), 364–370. doi: 10.1016/j.apmr.2007.08.139

Sassi, R. H., Dardouri, W., Yahmed, M. H., Gmada, N., Mahfoudhi, M. E., & Gharbi, Z. (2009). Relative and absolute reliability of a modified agility t-test and its relationship with vertical jump and straight sprint. *J Strength Cond Res*, *23*(6), 1644. doi: 10.1519/JSC.0b013e3181b425d2

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychol Rev*, 82(4), 225-260.

Semenick, D. (1990). Tests and measurements: *the T-test. Strength Cond J*, 12(1), 36-37.

Shadmehr, R., & Mussa-Ivaldi, S. (2012). *Biological learning and control: how the brain builds representations, predicts events, and makes decisions*. The MIT Press. doi: 10.7551/mitpress/9780262016964.001.0001

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annu Rev Neurosci*, *33*(1), 89-108. doi: 10.1146/annurev-neuro-060909-153135

Sheppard, J. M., & Young, W. B. (2006). Agility literature review: Classifications, training and testing. *J Sport Sci*, *24*(9), 919–932. doi: 10.1080/02640410500457109

Stevens, C. J., & Dascombe, B. J. (2015). The reliability and validity of protocols for the assessment of endurance sports performance: An updated review. *Measurement in Physical Education and Exercise Science*, *19*(4), 177–185. doi: 10.1080/1091367X.2015.1062381

Stuhr, C., Hughes, C. M. L., & Stöckel, T. (2018). Task-specific and variability-driven activation of cognitive control processes during motor performance. *Sci Rep*, *8*(1), 10811. doi: 10.1038/s41598-018-29007-3

Tahiroglu, F., & Shahidi, S. H. (2025). The effectiveness of sprint athletes in removing lactate after reaching maximum effort. Medicina dello Sport, 78(1), 70–78. doi: 10.1728/4552.45511

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. doi: 10.1126/science.7569931

Ye, J., Cheng, R., Gao, B., Wang, Y., & Wang, Y. (2025). Scientific assessment of agility performance in competitive sports: Evolution, application, reliability, and validity. *Strength Cond J, 47*(1), 56-75. doi: 10.1519/SSC.0000000000000862

Young, W., Rayner, R., & Talpey, S. (2021). It's time to change direction on agility research: a call to action. *Sports Med Open, 7*(1), 12. doi: 10.1186/s40798-021-00304-y