European Journal of Sport Science

WILEY

Official ECSS Journal
Open Access

ORIGINAL PAPER OPEN ACCESS

Validity and Inter-Device Reliability of an Artificial Intelligence App for Real-Time Assessment of 505 Change of Direction Tests

Francisco J. Barrera-Domínguez[1] | Paul A. Jones[2] | Bartolomé J. Almagro[1] | Jorge Molina-López[1]

[1]Faculty of Education, Psychology and Sport Sciences, COIDESO, University of Huelva, Huelva, Spain | [2]Directorate of Sport, Exercise and Physiotherapy, University of Salford, Salford, UK

Correspondence: Francisco J. Barrera-Domínguez (francisco.barrera@ddi.uhu.es)

Received: 2 October 2024 | **Revised:** 18 December 2024 | **Accepted:** 6 January 2025

Funding: This paper is part of the first author's doctoral thesis carried out in the Doctoral Programme of the University of Huelva (Spain), thanks to the support and funding of the Formación del Profesorado Universitario Programme (FPU22/01057), run by the Ministerio de Ciencias, Innovación y Universidades, Government of Spain.

Keywords: agility | computer vision | motion capture | multidirectional speed | technology

ABSTRACT

The present study aimed to explore the validity and inter-device reliability of a novel artificial intelligence app (Asstrapp) for real-time measurement of the traditional (tra505) and modified-505 (mod505) change of direction (COD) tests. Twenty-five male Sports Science students (age, 23.5 ± 3.27 years; body height, 178 ± 9.76 cm; body mass, 79.4 ± 14.7 kg) completed 12 trials each, consisting of six tra505 and six mod505 trials. Completion times were simultaneously recorded via single-beam electronic timing gates (ETG) and two different iPhones (APP1 and APP2). In total 300 trials were collected across the two tests, using all three devices, to establish the reliability and validity of the app. The coefficient of variation indicated a similar level of dispersion between the ETG (≤ 2.73%), APP1 (≤ 2.39%) and APP2 (≤ 2.52%). Intraclass correlation coefficients (ICC) revealed excellent reliability among the three timing devices (ICC ≥ 0.99) and Asstrapp relative reliability was excellent for both APP1 (ICC ≥ 0.91) and APP2 (ICC ≥ 0.91). There was a practically perfect correlation and agreement between ETG and Asstrapp (APP1: $r = 0.97$; APP2: $r = 0.97$) for both COD tests. However, small but significant differences were found between smartphones and ETG for tra505 (ES ≤ 0.33; $p < 0.05$). Collectively, these findings support the use of Asstrapp for real-time assessment of both 505 COD tests.

1 | Introduction

Changes of direction (COD) are key performance metrics in many field and court sports and are often decisive in key moments in a match, such as scoring a goal (Taylor et al. 2017). Previous studies have revealed that professional football players execute over 726 cuts during a match (Bloomfield, Polman and O'Donoghue 2007). Moreover, most movements preceding a goal situation involve decelerations and turns (Martínez-Hernández, Quinn and Jones 2023). In basketball, players typically perform a COD every 1–3 s within a game (Klusemann et al. 2013), being the action that is most frequently performed regardless of playing position or gender (Salazar, Castellano and Svilar 2020). This relevance of COD is also evident in handball, where more than 60% of the maximum intensity actions involve a COD (Povoas et al. 2012). Beyond team sports, COD actions also play an important role in racket sports (Schneider, Rothschild and Uthoff 2023). With COD being such a decisive

Summary

- Asstrapp is a mobile application specifically designed to measure the total time in sporting tests automatically and was able to provide a valid and reliable measurement in both 505 COD tests compared to the gold-standard.
- Asstrapp demonstrated perfect validity and agreement in inter-smartphone measurements in both COD tests, tra505 and mod505.
- It is important to consider the characteristics of the COD test when selecting a measurement device. Particularly for tests involving a flying start, such as the tra505 test, where caution should be exercised when using single-beam ETG.

physical quality in competition across a wide range of sports as it has been noted, it is not surprising that the COD assessment is frequently included in the sports testing battery that sport scientists and coaches perform on their athletes (Nimphius et al. 2018).

The assessment of COD has been extensively studied, resulting in the development of over 48 different tests to analyse COD performance (Sugiyama et al. 2021). Nevertheless, the most commonly used test is the traditional (tra505, flying start) or modified 505 (mod505, stationary start) COD test with a 180-degree cut (Ryan et al. 2021b). Additionally, a multitude of technological tools have been used in scientific literature to measure performance in these COD tests, with electronic timing gates (ETG) being widely accepted as time-based gold standard instruments (Altmann et al. 2018). However, this technology has some drawbacks, such as its high cost, which may prevent its accessibility to coaches and institutions with budget constraints, making it difficult to use in athletes' daily practice. As a more affordable and portable alternative, coaches may resort to using a stopwatch in daily practice. Nevertheless, previous research has indicated that manual timing exhibits large absolute errors during linear sprinting and COD tasks (Chen et al. 2021), so its use has not been recommended for the assessment of these maximum speed actions.

In order to overcome all these aforementioned limitations, a plethora of smartphone applications (apps) have appeared as a cost-effective and portable alternative to traditional laboratory equipment for assessing sports performance. Although not all apps have undergone empirical tests of validity and reliability, previous research has shown that these tools can be valid and reliable for assessing physical capacities such as jumping actions, barbell velocity in strength exercises and speed in sprint and COD tests (Pérez-Castilla et al. 2019; Romero-Franco et al. 2017; Silva et al. 2021; Turan et al. 2022). This reliability is attributed to the high-speed cameras integrated in current smartphones. However, these apps have typically lacked real-time feedback, requiring coaches to manually assess their athletes afterwards, resulting in time-consuming evaluations that cannot facilitate real-time decision-making in sports training. Advances in artificial intelligence (AI) and computer vision offer the potential to overcome this limitation by enabling smartphones to autonomously identify measurement targets, analyse them and provide test results in real time. Due to the novelty of this technology, there are limited validation and reliability studies employing AI in sports testing. Nonetheless, recent research has validated AI apps for tasks such as tracking the bar path in weightlifting movements (Balsalobre-Fernández et al. 2020), assessing joint range of motion (Carrasco-Uribarren et al. 2023), measuring barbell velocity in strength exercises (Taber et al. 2023) or determining cricket bowling line and length (Tissera et al. 2024). However, to the best of the authors' knowledge, no studies have yet assessed the validity and reliability of any AI or computer vision app in COD tests.

Therefore, the main purpose of this research was to examine the validity and reliability of Asstrapp in the tra505 and mod505 COD tests. Additionally, each mobile device model has different features and characteristics that might affect the results of Asstrapp's measurements. Inter-smartphones validity and reliability were analysed by comparing Asstrapp on two different mobile phones. Based on previous literature that analysed the validity of AI apps to measure several physical capacities, it hypothesised that Asstrapp may be a valid and reliable alternative for the measurement of total time in the tra505 and mod505 when compared with ETG.

2 | Materials and Methods

2.1 | Design

A reliability and validity study were carried out following the guidelines for reliability studies (Kottner et al. 2011). Testing took place in the human performance laboratory on an indoor track (Mondo, SportsFlex, 10 mm; Mondo America, Inc., Mondo, Summit, NJ) in a single session. Prior to starting the tests, a 15-min warm-up was conducted; which began with a general activation including light intensity jogging, two sets of dynamic stretching (four knee hug-moving, four walking quad stretches, two inchworms and two world's greatest stretch on each side) and three sets of acceleration-deceleration drills, followed by three submaximal mod505 trials. After that, participants began to perform the tra505 and mod505. Moreover, each participant was instructed to attend the testing sessions adequately hydrated and rested, with no high intensity training in the previous 24 h and to control their caffeine and food intake for at least 3 h before evaluation.

Each participant completed a total of 12 attempts. First, six mod505 and then, six tra505 trials. Each trial was interspersed with 3 min of rest. All attempts were timed simultaneously in three ways: using sets of single-beam ETG; and on the other hand, using two different iPhones. Times were recorded to the nearest 0.001 s. A total of 300 trials were collected with each tool in both COD tests for further analysis.

2.2 | Participants

The sample size was calculated based on an effect size of 0.5, an alpha level of 0.05 and power value of 0.80 (Faul et al. 2007) using G*Power software (version 3.1.9.6, Kiel, Germany). Twenty-five male Sports Science students (age, 23.5 ± 3.27 years; body height, 178 ± 9.76 cm; body mass, 79.4 ± 14.7 kg) from the School of Health & Society at the University of Salford were selected to voluntarily participate in this study. All participants met the following inclusion criteria: at least 1 year of sport training experience, no injuries in the last 6 months, no health problems, practice a sport besides the sports science degree at university and age ranged from 20 to 30 years. All of them were previously informed of the possible risks and benefits of participating in the study and, before the beginning of the testing, they gave their written consent. This research was approved by the Andalusian Biomedical Research Ethics Committee (reference number: TES_COD_23) and University of Salford Ethics Committee (reference number: HSR2324_001) in accordance with the guidelines established in the Declaration of Helsinki.

2.3 | Procedures

Traditional and Modified 505 COD Setup. A schematic of the testing set-up is provided in Figure 1. Participants performed six mod505 trials and six tra505 trials. Each test was performed on both sides (first three right leg trials and then, three left leg trials for each test). The starting position was standardized for all participants. Each participant was positioned 0.5 m behind the start line, in a two-point split stance, to try to avoid early triggering of the timing gates. For the tra505 and mod505, the starting line was 15 and 5 m from the turning point, respectively (Figure 1). Both tests seek to assess an individual's ability to sprint 5 m, perform a 180° COD and re-accelerate 5 m to the finish line; but the trad505 has 10 m flying start and the mod505 has a standing start just behind the start line (Ryan et al. 2021a). If the participant slid or turned prematurely the trial was discarded and subsequently another trial was performed after 3 min of rest.

Electronic Timing Gate (ETG) device. A set of single-beam Brower ETG (Draper, UT) at a frequency of 480 Hz were used as the gold standard to measure the execution time of the attempts. The ETC was placed at the start/finish line, 2 m apart from each other and at a height of 1 m (approximately the height of the participants' hips) to ensure only one body part breaks the beam. The timing cells system can send radio transmissions up to 250 m and is accurate to the thousandth of a second, making it a highly precise timing tool (Shalfawi et al. 2014).

Asstrapp Smartphone App. Two iPhones, one basic and one pro, were used to analyse Asstrapp's reliability. The app (Asstrapp, Version 3.0) was installed on an iPhone 11 (APP1) and an iPhone 15 Pro (APP2) both running iOS 17.1 (Apple Inc., Cupertino, USA). Videos were captured with the on-board camera of the iPhones using Asstrapp, recording at HD 1080p resolution and 60 Hz sample rate; so, the software automatically analysed them and provided real-time feedback. Each iPhone was placed in a landscape orientation on a tripod at 1 m high and 4 m away from the start/finish line, perpendicular to the direction of movement of the test (Figure 1). In this way it ensures that the participant's full body is focussed within the Asstrapp measurement screen as it passes through the start/finish line. The centre of the Asstrapp measurement screen must coincide precisely with the start/finish line, so that the software automatically detects the start/finish of the test when the participant's hip crosses the start/finish line.

The app was developed by using Apple's open-source Swift 5 programming language in Xcode 16 for macOS (Apple Inc., Cupertino, USA). Live object recognition features were included using Apple's computer-vision framework Vision 2.0 (Apple Inc., Cupertino, USA) for real-time detection and tracking of humans. Specific, technical details of this framework can be found following this link: https://developer.apple.com/documentation/vision. This computer vision system is based on a pre-trained model, provided by Apple (Apple Inc., Cupertino, USA), to detect and track multiple body points, facilitating accurate tracking of the person's movement throughout the test. Specifically, the framework was integrated into Asstrapp to detect the person's hip point, so that it can detect when the person crosses the start/finish line of the test without the movement of the limbs influencing the test result. Additionally, the app automatically stores the collected data on a server by calling an Application Programming Interface built in Hypertext Pre-Processor.

2.4 | Statistical Analyses

Means ± standard deviations (SD) were used to describe variables. Normality and homogeneity were verified using Shapiro-Wilk and Levene's tests, respectively. The intraclass correlation coefficient (ICC, two-way random, absolute agreement) with 95% confidence intervals (CI) was used to analyse the within session relative reliability of the total time measures in the COD tests using each device and mobile phone app. ICCs were also used to explore the agreement between ETG and mobile phone app. ICCs were interpreted as follow: ICC < 0.50 = 'poor', 0.50–0.75 = 'moderate', 0.75–0.90 = 'good', > 0.90 = 'excellent' (Koo and Li 2016). The coefficient of variation (CV) was calculated as: (SD/average) × 100 and used to analyse the stability of timing systems, with a CV < 5% considered as acceptable for the absolute within-session reliability of each device (Atkinson and Nevill 1998). A linear regression with Pearson's (r) correlation coefficient with 95% CI, the standard error of the estimate (SEE) and the correction equations were analysed to assess the concurrent validity of Asstrapp, in comparison with the ETG. Also, the possibility of collinearity between devices was tested by the variance inflation factor (VIF; VIF < 10), tolerance (> 0.2) and verified by the Durbin–Watson test. The strength of the r coefficients was interpreted as follows: $r < 0.10$ = 'trivial', 0.10–0.29 = 'small', 0.30–0.49 = 'moderate', 0.50–0.69 = 'high', 0.70–0.89 = 'very high' and > 0.90 = 'practically perfect' (Hopkins et al. 2009). Paired samples t-test and Bland–Altman plots were used to identify potential systematic bias, reported via mean bias, standard deviations and the analysis of the regression line on the Bland-Altman plots (Bland and Altman 1986). Cohen's d was used to assess the mean differences between the measures obtained with each instrument; thresholds for qualitative descriptors of Cohen's d were set at, < 0.20 'trivial', 0.20–0.50 'small', 0.50–0.80 'moderate' and > 0.80 'large' (Cohen 1988). Statistical significance was set at $p < 0.05$. All statistical analysis was performed using Jamovi 2.3.21 for macOS.

3 | Results

Reliability statistics of the ETG, APP1 and APP2 is shown in Table 1. Both tra505 and mod505 demonstrated highly acceptable levels of stability across different timing systems (tra505 CV: ETG = 2.23%; APP1 = 2.15%; and APP2 = 2.13%; and mod505 CV: ETG = 2.73%; APP1 = 2.39%; and APP2 = 2.52%). No significant differences were detected in CV values between trad505 trials (ES $\leq$ 0.089; $p \geq 0.631$) and mod505 trials (ES $\leq$ 0.171; $p \geq 0.356$) measured with ETG, APP1 and APP2. The reproducibility of data provided by each measuring devices for the assessment of trad505 and mod505 showed good to excellent within-device agreement (Table 1). Additionally, excellent agreement was observed between ETG, APP1 and APP2 for the measurement of total time in tra505 COD tests (ICC = 0.987 [95% CI = 0.982–0.990]) and mod505 (ICC = 0.993

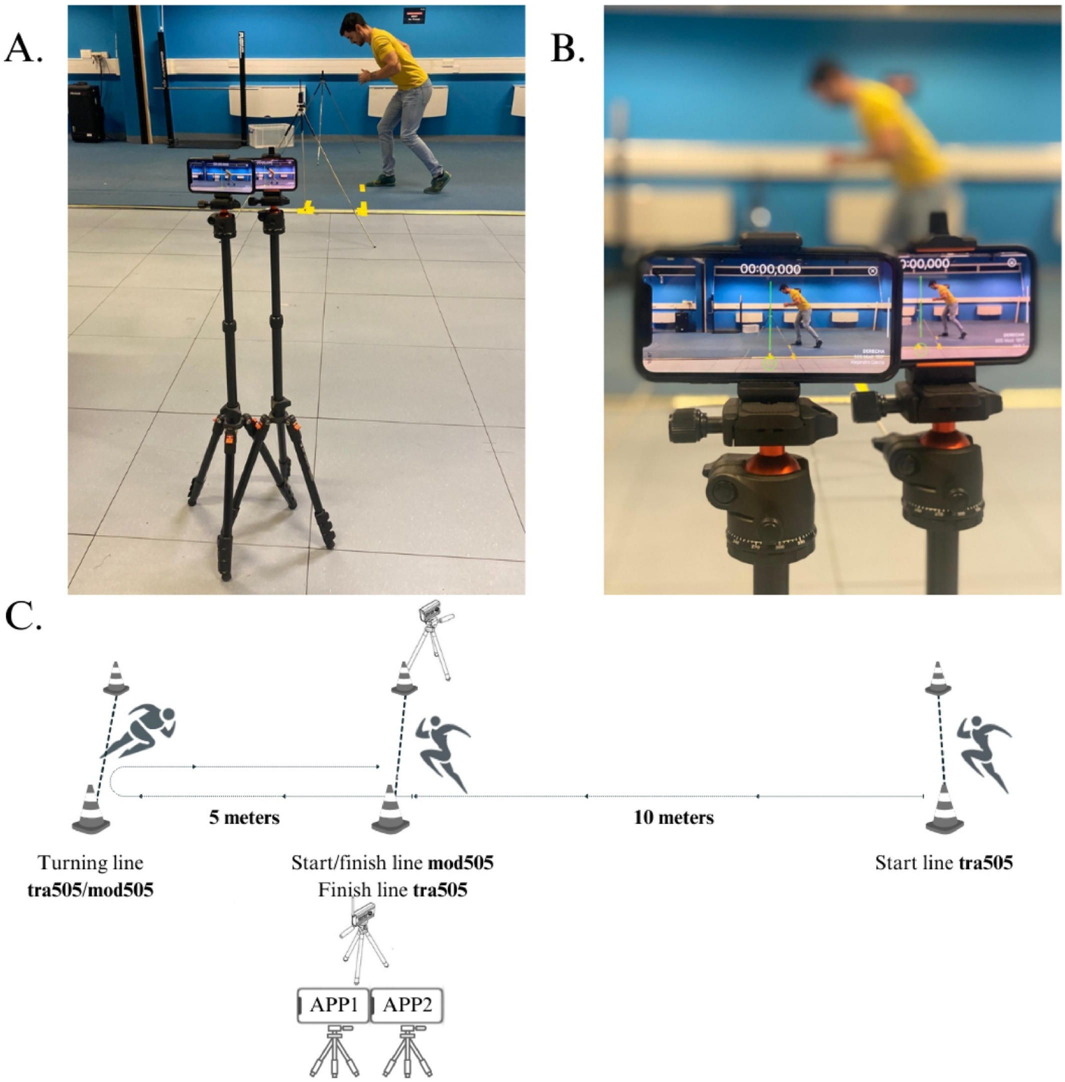

FIGURE 1 | Illustration of the tra505 and mod505 COD tests, showing where the ETG, APP1 and APP2 were placed. (A) Overall assessment set-up in the human performance lab, with the ETG and Asstrapp AI app on two different iPhones; (B) Fitting of the Asstrapp measurement screen with test start/finish line; and (C) Schematic diagram of the tra505 and mod505 COD tests set-up.

TABLE 1 | Two-way random intraclass correlation coefficient and coefficient of variation with 95% confidence intervals assessing within session reliability of each device.

	Tra505		Mod505	
	ICC (95% CI)	**CV (95% CI)**	**ICC (95% CI)**	**CV (95% CI)**
ETC	0.893 (0.811; 0.944)	2.229 (1.597; 2.861)	0.920 (0.858; 0.958)	2.729 (2.018; 3.440)
APP1	0.919 (0.856; 0.957)	2.151 (1.578; 2.724)	0.933 (0.882; 0.965)	2.387 (1.769; 3.005)
APP2	0.918 (0.856; 0.957)	2.135 (1.585; 2683)	0.925 (0.867; 0.960)	2.520 (1.864; 3.176)

Abbreviations: APP1, Asstrapp on an iPhone 11; APP2, Asstrapp on an iPhone 15 Pro; CI, confidence intervals; CV, coefficient of variation; ETC, electronic timing gate; ICC, intraclass correlation coefficient; Mod505, modified 505 change of direction test; Tra505, traditional 505 change of direction test.

[95% CI = 0.991–0.004]). Moreover, Asstrapp showed excellent inter-smartphones' reliability for tra505 (ICC = 0.971 [95% CI = 0.955–0.984]) and mod505 (ICC = 0.978 [95% CI = 0.965–0.999]).

Figure 2 shows scatter plot graphs of the linear regression analysis with Pearson's *r* correlation for the trad505 and mod505 tests comparing between measuring devices. Asstrapp demonstrated practically perfect validity in comparison with ETG for mod505

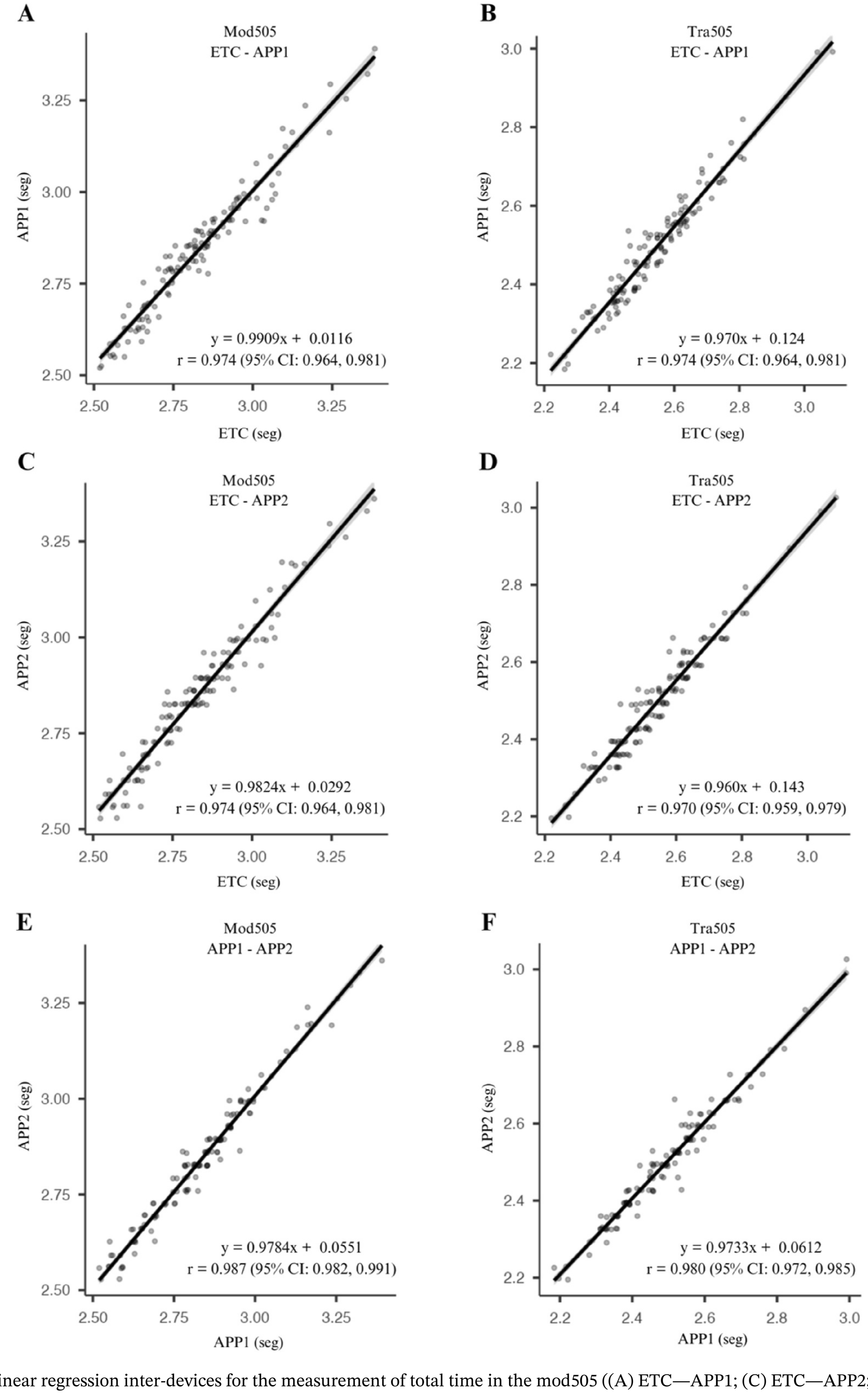

FIGURE 2 | Linear regression inter-devices for the measurement of total time in the mod505 ((A) ETC—APP1; (C) ETC—APP2; (E) APP1—APP2) and tra505 ((B) ETC—APP1; (D) ETC—APP2; (F) APP1—APP2) COD tests.

(APP1: $r = 0.974$; SEE = 0.019s; and correction equation = $0.991x + 0.012$; $p < 0.001$; and APP2: $r = 0.974$; SEE = 0.019s; and correction equation = $0.982x + 0.029$; $p < 0.001$); and for trad505 (APP1: $r = 0.974$; SEE = 0.018s; and correction equation = $0.970x + 0.124$; $p < 0.001$; and APP2: $r = 0.970$; SEE = 0.020s; and correction equation = $0.960x + 0.143$; $p < 0.001$). Furthermore, it also showed a perfect validity inter-smartphone when comparing the results for mod505 ($r = 0.987$; SEE = 0.013s; and correction equation = $0.978x + 0.055$; $p < 0.001$); and for trad505 ($r = 0.980$; SEE = 0.016s; and correction equation = $0.973x + 0.061$; $p < 0.001$). No collinearity was observed in the linear regression analysis (tolerance ≥ 0.98, VIF ≤ 1.05 and $d \geq 2.02$ for both COD tests).

The comparison of mean results between devices is presented in Table 2. Non-significant differences were found between devices in the mod505 tests and the pooled data from the COD tests evaluated (ES ≤ 0.117; $p \geq 0.320$). However, tra505 showed small significant differences between smartphones and ETG (ES ≤ 0.329; $p < 0.014$), although no significant differences were observed in tra505 between smartphones devices using Asstrapp (ES = 0.035; $p = 0.764$).

Figure 3 shows Bland–Altman plots to detect potential systematic bias. Mean bias estimate for mod505 between ETG-APP1 was −0.014s (95% CI: −0.021, −0.008s); ETG-APP2 was −0.021s (95% CI: −0.027, −0.014s); and APP1-APP2 was −0.006s (95% CI: −0.011, −0.002s). For trad505 the mean bias estimate between ETG-APP1 was 0.049s (95% CI: 0.044, 0.055s); ETG-APP2 was 0.044s (95% CI: 0.038, 0.051s); and between APP1-APP2 was −0.005s (95% CI: −0.10, −0.001s). These results indicate perfect agreement between all instruments across different COD tests.

4 | Discussion

The main purpose of the present study was to examine the validity and reliability of a novel AI app in the tra505 and mod505 COD tests. Additionally, inter-smartphones validity and reliability were analysed by comparing Asstrapp on two different mobiles. Results showed that Asstrapp is a highly valid and reliable tool for measuring tra505 and mod505 COD tests. Thus, Asstrapp could be a very cost-effective and automatic alternative for COD performance assessment. This could make COD performance assessment quicker and more accessible for strength and conditioning coaches and trainers.

Although the use of laboratory instruments is highly sensitive, accurate and considered gold-standard, the use of mobile apps in sports testing is currently on the rise and has many advantages such as being affordable for all professionals, as well as their easy use and portability, among others. This makes them increasingly useful for sports researchers and physical trainers if they have been shown to be valid and reliable (Peart, Balsalobre-Fernández and Shaw 2019). From a reliability standpoint, the results of the present study showed that Asstrapp exhibits excellent relative within-session reliability (APP1, ICC ≥ 0.911; APP2, ICC ≥ 0.905; Table 1) when measuring execution time in both COD tests analysed. Moreover, the ICC was

TABLE 2 | Comparison (mean ± SD) and Cohen's d effect size with 95% confidence intervals denoting practical differences between measurement devices in tra505, mod505 and pooled data.

	ETC	APP1	APP2	ETC—APP1			ETC—APP2			APP1—APP2		
	Mean ± SD	Mean ± SD	Mean ± SD	p	ES (95% CI)	Descriptor	p	ES (95% CI)	Descriptor	p	ES (95% CI)	Descriptor
Tra505 (s)	2.539 ± 0.15	2.489 ± 0.15	2.491 ± 0.15	0.006	0.329 (0.096; 0.561)	Small	0.014	0.292 (0.059; 0.524)	Small	0.764	−0.035 (−0.266; 0.196)	Trivial
Mod505 (s)	2.831 ± 0.17	2.845 ± 0.17	2.852 ± 0.17	0.493	−0.081 (−0.312; 0.150)	Trivial	0.320	−0.117 (−0.349; 0.114)	Trivial	0.753	−0.037 (−0.268; 0.194)	Trivial
Pooled (s)	2.685 ± 0.22	2.667 ± 0.24	2.673 ± 0.24	0.356	0.077 (−0.086; 0.240)	Trivial	0.541	0.051 (−0.112; 0.214)	Trivial	0.768	−0.025 (−0.188; 0.139)	Trivial

Abbreviations: APP1, Asstrapp on an iPhone 11; APP2, Asstrapp on an iPhone 15 Pro; CI, confidence intervals; ES, effect size; ETC, electronic timing gate; Mod505, modified 505 change of direction test; s, seconds; SD, standard deviation; Tra505, tradicional 505 change of direction test.

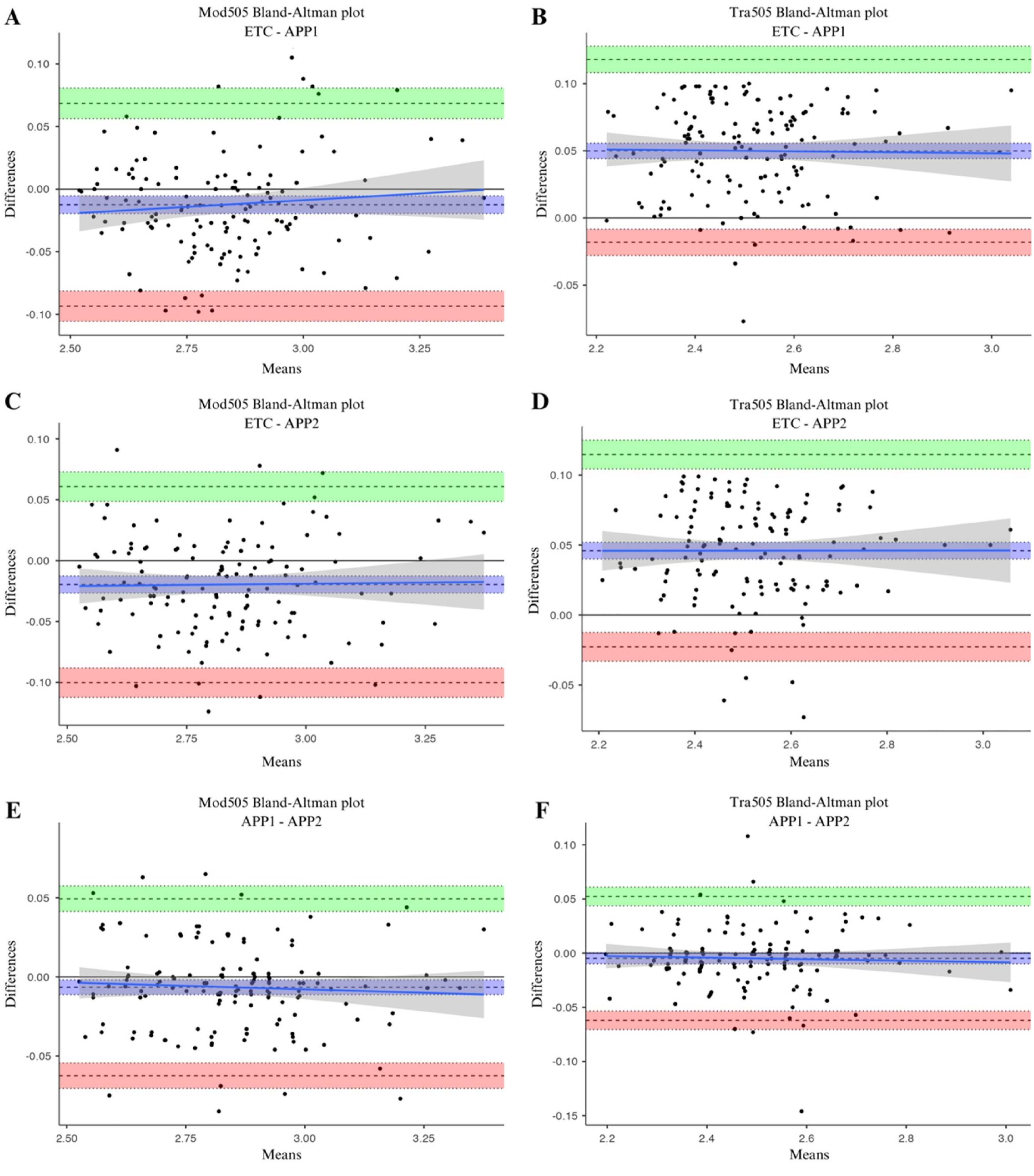

FIGURE 3 | Bland–Altman plot showing levels of agreement in the measurement inter-devices for mod505 ((A) ETC—APP1; (C) ETC—APP2; (E) APP1—APP2) and trad505 ((B) ETC—APP1; (D) ETC—APP2; (F) APP1—APP2) COD test. Including the mean bias estimate bias (with 95% CI), both lower and upper limits of agreement (with 95% CI) and the regression line of the residual (blue line with 95% CI).

also used to compare agreement between the ETG and Asstrapp showing almost perfect inter-devices reliability (ICC ≥ 0.987). These ICC values are higher than those found for other apps in previous studies (ICC = 0.671–0.840) (Balsalobre-Fernández et al. 2019). Additionally, the absolute reliability values assessed through the CV showed that the results of all measuring tools were acceptable and stable, with the CV being very similar to inter-devices and smartphones in the different COD tests (Table 1). The CV found are in line with previous studies analysing the reliability of a video-based app assessing mod505 (CV = 3.2%) (Balsalobre-Fernández et al. 2019) and tra505 (CV = 6.3%) (Chen et al. 2021). Therefore, professionals and Asstrapp users may be confident that the results provided by the app to measure the execution time of COD tests are reliable.

Due to the growing use of apps in sports testing, previous research has already shown that current smartphone technology makes it possible to validate the execution time in COD tests (r = 0.964; 95% CI = 0.95–1.00) using slow-motion video-based apps (Balsalobre-Fernández et al. 2019). However, scientifically validated apps to assess COD performance rely on external observers so far and although these apps have been tested for inter-observer validity (Chen et al. 2021), the measurement is not done automatically. So, no instant feedback is provided and professionals must invest significant time in video analysis. The app analysed in the present study uses AI to automatically give the COD test result. With the help of a machine learning model that has been programmed to detect people in motion, the AI-based smartphone app itself is able to detect when a person crosses the start and finish line of a COD test. This allows Asstrapp to do the same task as an ETG by calculating the time it takes for a participant to complete any COD test automatically and providing real-time feedback. The current research results reveal a very high concurrent validity of Asstrapp compared to ETG, showing an almost perfect association inter-devices ($r \geq 0.970$; $r^2 \geq 0.942$) and the slope coefficient was very close to the identity line (slope ≥ 0.960) for both COD tests assessed. Furthermore, a practically perfect validity was also observed in inter-smartphones using Asstrapp, showing an even stronger association ($r \geq 0.980$; $r2 \geq 0.960$). This agreement in the results reported inter-devices and inter-smartphones was also supported by Bland-Altman plots. Simply put, the values measured for both inter-devices and inter-smartphones were very consistent in both COD tests assessed, confirming the validity of Asstrapp. On the other hand, including measures from the same participant in a regression could lead to collinearity of the data, producing an overestimation of the model (Naclerio and Larumbe-Zabala 2017). But even though six attempts from each participant were included, this did not affect the fit of the linear regression model as no collinearity was observed in the sample data.

Finally, although all previous analysis denotes a very high validity of Asstrapp to measure execution time in COD tests, it is recommended to be cautious with the measurement instrument used according to the COD test assessed. Because small significant differences in tra505 execution time were observed between ETG and Asstrapp on both mobiles. These differences inter-devices were only found for the trad505 and no differences were found for the mod505 test and the pooled results. This finding may be explained by several reasons. Firstly, the sampling frequency inter-devices was different. Furthermore, the ETC used could be activated by cutting the single-beam through either limb of the athlete while running, while Asstrapp's AI is programmed so that only the person's hip activates the system when it crosses the start and finish test line to avoid any early start-finish cut-off, errors in the measurement and providing very similar inter-smartphone results. These differences in sampling frequency and tool activation methods may explain the difference observed just in tra505, which is a test with a flying start and higher speeds. Although the ETG was adjusted as recommended by previous scientific literature (Haugen, Tønnessen and Seiler 2015), showed a longer execution time for each tra505 trial (+0.049 vs. APP1; +0.044 vs. APP2). However, results showed stable inter-device differences in all tra505 trials. These differences were similar to those found in previous studies (Haugen et al. 2014) comparing a single-beam versus a dual-beam ETG system (+0.03 to +0.05 vs. dual-beam ETG). This could be explained because the activation method of the Asstrapp timing system is likely to be more similar to the dual-beam ETG system. These previous reasons could explain the small differences observed between Asstrapp and single-beam ETG in the tra505, whereby the result provided by Asstrapp's AI may be the one that best fits the reality.

Despite the novelty and usefulness of the findings for strength and conditioning coaches, there are some limitations in this study that should be acknowledged. Mainly, as the present research has highlighted, the validity data of Asstrapp could vary depending on the COD test analysed. Although there are many COD tests, only mod505 and tra505 have been analysed in this study, so caution is advised when using the app to assess other COD tests. Secondly, it should be noted that Asstrapp is currently only available for iOS and although this validation study has been conducted with two different iOS models to analyse validity and reliability inter-smartphone, it would recommend developing an equivalent AI software for smartphones with Android operating system, thus encouraging the COD assessment to all sport professionals.

5 | Conclusion

Accuracy and stability of measurement are indispensable characteristics for timing devices in COD testing. The findings of the present study showed that the Asstrapp is a highly valid and reliable tool for measuring the execution time of the mod505 and tra505 COD tests. Moreover, Asstrapp is an affordable and accessible sports testing instrument for all coaches, a user-friendly and quick-to-use tool, it is more portable than traditional laboratory instruments and has the advantage that it offers immediate feedback. For all these reasons, this app is recommended to any sports professional to assess, quantify and train COD.

Author Contributions

Research concept and study design: F.J.B.-D. and J.M.-L. Literature review: F.J.B.-D. and P.A.J. Data collection: F.J.B.-D. and P.A.J. Data analysis and interpretation: F.J.B.-D., B.J.A. and J.M.-L. Statistical analyses: F.J.B.-D., B.J.A. and J.M.-L. Writing of the manuscript: F.J.B.-D. and P.A.J. and Reviewing/editing a draft of the manuscript: P.A.J., B.J. A. and J.M.-L.

Acknowledgements

The authors are particularly grateful to all the students who voluntarily participated in the present study. Authors are also grateful for the support of the 'Network of Sports Functional Dynamometry' (09/UPB/23) and the Centro de Investigación en Pensamiento Contemporáneo e Innovación para el Desarrollo Social (COIDESO) of the University of Huelva (Spain). This paper is part of the first author's doctoral thesis carried out in the Doctoral Programme of the University of Huelva (Spain), thanks to the support and funding of the Formación del Profesorado Universitario Programme (FPU22/01057), run by the Ministerio de Ciencias, Innovación y Universidades, Government of Spain. Funding for open access charge: University of Huelva/CBUA.

Ethics Statement

This study was conducted following the principles of the Declaration of Helsinki and approved by the Andalusian Biomedical Research Ethics Committee (reference number: TES_COD_23) and University of Salford ethics committee (reference number: HSR2324-001).

Conflicts of Interest

The first author of the present investigation is the creator of the app mentioned. To guarantee data independency, data were collected and analysed by independent researchers not related with the app's development (specifically, the second, third and last authors of the present manuscript).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, F.J.B.-D., upon reasonable request.

References

Altmann, S., S. Ringhof, B. Becker, A. Woll and R. Neumann. 2018. "Error-Correction Processing in Timing Lights for Measuring Sprint Performance: Does It Work?" *International Journal of Sports Physiology and Performance* 13, no. 10: 1400–1402. https://doi.org/10.1123/IJSPP.2017-0596.

Atkinson, G. and A. M. Nevill. 1998. "Statistical Methods for Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine." *Sports Medicine* 26, no. 4: 217–238. https://doi.org/10.2165/00007256-199826040-00002.

Balsalobre-Fernández, C., C. Bishop, J. V. Beltrán-Garrido, et al. 2019. "The Validity and Reliability of a Novel App for the Measurement of Change of Direction Performance." *Journal of Sports Sciences* 37, no. 21: 2420–2424. https://doi.org/10.1080/02640414.2019.1640029.

Balsalobre-Fernández, C., G. Geiser, J. Krzyszkowski and K. Kipp. 2020. "Validity and Reliability of a Computer-Vision-Based Smartphone App for Measuring Barbell Trajectory During the Snatch." *Journal of Sports Sciences* 38, no. 6: 710–716. https://doi.org/10.1080/02640414.2020.1729453.

Bland, J. M. and D. G. Altman. 1986. "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement." *Lancet* 1: 307–310.

Bloomfield, J., R. Polman and P. O'Donoghue. 2007. "Physical Demands of Different Positions in FA Premier League Soccer." *Journal of Sports Science and Medicine* 6, no. 1: 63.

Carrasco-Uribarren, A., X. Marimon, F. Dantony, et al. 2023. "A Computer Vision-Based Application for the Assessment of Head Posture: A Validation and Reliability Study." *Applied Sciences (Switzerland)* 13, no. 6: 3910. https://doi.org/10.3390/APP13063910.

Chen, Z., C. Bian, K. Liao, C. Bishop and Y. Li. 2021. "Validity and Reliability of a Phone App and Stopwatch for the Measurement of 505 Change of Direction Performance: A Test-Retest Study Design." *Frontiers in Physiology* 12. https://doi.org/10.3389/FPHYS.2021.743800.

Cohen, J. 1988. "Differences Between Correlation Coefficients." In *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. New York: Routledge. https://doi.org/10.4324/9780203771587.

Faul, F., E. Erdfelder, A.-G. Lang and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191. https://doi.org/10.3758/BF03193146.

Haugen, T., E. Tønnessen and S. Seiler. 2015. "Correction Factors for Photocell Sprint Timing With Flying Start." *International Journal of Sports Physiology and Performance* 10, no. 8: 1055–1057. https://doi.org/10.1123/ijspp.2014-0415.

Haugen, T. A., E. Tønnessen, I. S. Svendsen and S. Seiler. 2014. "Sprint Time Differences Between Single- and Dual-Beam Timing Systems." *Journal of Strength and Conditioning Research* 28, no. 8: 2376–2379. https://doi.org/10.1519/JSC.0000000000000415.

Hopkins, W. G., S. W. Marshall, A. M. Batterham and J. Hanin. 2009. "Progressive Statistics for Studies in Sports Medicine and Exercise Science." *Medicine & Science in Sports & Exercise* 41, no. 1: 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278.

Klusemann, M. J., D. B. Pyne, W. G. Hopkins and E. J. Drinkwater. 2013. "Activity Profiles and Demands of Seasonal and Tournament Basketball Competition." *International Journal of Sports Physiology and Performance* 8, no. 6: 623–629. https://doi.org/10.1123/IJSPP.8.6.623.

Koo, T. K. and M. Y. Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15, no. 2: 155–163. https://doi.org/10.1016/J.JCM.2016.02.012.

Kottner, J., L. Audigé, S. Brorson, et al. 2011. "Guidelines for Reporting Reliability and Agreement Studies (GRRAS) Were Proposed." *Journal of Clinical Epidemiology* 64, no. 1: 96–106. https://doi.org/10.1016/J.JCLINEPI.2010.03.002.

Martínez-Hernández, D., M. Quinn and P. Jones. 2023. "Most Common Movements Preceding Goal Scoring Situations in Female Professional Soccer." *Science and Medicine in Football* 8, no. 3: 260–268. https://doi.org/10.1080/24733938.2023.2214106.

Naclerio, F. and E. Larumbe-Zabala. 2017. "Technical Note on Using the Movement Velocity to Estimate the Relative Load in Resistance Exercises - Letter to the Editor." *Sports Medicine International Open* 2, no. 1: E16. https://doi.org/10.1055/S-0043-118710.

Nimphius, S., S. J. Callaghan, N. E. Bezodis and R. G. Lockie. 2018. "Change of Direction and Agility Tests: Challenging Our Current Measures of Performance." *Strength and Conditioning Journal* 40, no. 1: 26–38. https://doi.org/10.1519/SSC.0000000000000309.

Peart, D. J., C. Balsalobre-Fernández and M. P. Shaw. 2019. "Use of Mobile Applications to Collect Data in Sport, Health and Exercise Science: A Narrative Review." *Journal of Strength and Conditioning Research* 33, no. 4: 1167–1177. https://doi.org/10.1519/JSC.0000000000002344.

Pérez-Castilla, A., A. Piepoli, G. Delgado-García, G. Garrido-Blanca and A. García-Ramos. 2019. "Reliability and Concurrent Validity of Seven Commercially Available Devices for the Assessment of Movement Velocity at Different Intensities During the Bench Press." *Journal of Strength and Conditioning Research* 33, no. 5: 1258–1265. https://doi.org/10.1519/JSC.0000000000003118.

Póvoas, S. C. A., A. F. T. Seabra, A. A. M. R. Ascensão, J. Magalhães, J. M. C. Soares and A. N. C. Rebelo. 2012. "Physical and Physiological Demands of Elite Team Handball." *Journal of Strength and Conditioning Research* 26, no. 12: 3365–3375. https://doi.org/10.1519/JSC.0B013E318248AEEE.

Romero-Franco, N., P. Jiménez-Reyes, A. Castaño-Zambudio, et al. 2017. "Sprint Performance and Mechanical Outputs Computed With an iPhone App: Comparison With Existing Reference Methods." *European Journal of Sport Science* 17, no. 4: 386–392. https://doi.org/10.1080/17461391.2016.1249031.

Ryan, C., A. Uthoff, C. McKenzie and J. Cronin. 2021a. "Traditional and Modified 5-0-5 Change of Direction Test." *Strength and Conditioning Journal* 44, no. 4: 22–37. https://doi.org/10.1519/SSC.0000000000000691.

Ryan, C., A. Uthoff, C. McKenzie and J. Cronin. 2021b. "Traditional and Modified 5-0-5 Change of Direction Test: Normative and Reliability Analysis." *Strength and Conditioning Journal* 44, no. 4: 22–37. https://doi.org/10.1519/SSC.0000000000000691.

Salazar, H., J. Castellano and L. Svilar. 2020. "Differences in External Load Variables Between Playing Positions in Elite Basketball Match-Play." *Journal of Human Kinetics* 75, no. 1: 257–266. https://doi.org/10.2478/HUKIN-2020-0054.

Schneider, C., J. Rothschild and A. Uthoff. 2023. "Change-of-Direction Speed Assessments and Testing Procedures in Tennis: A Systematic Review." *Journal of Strength and Conditioning Research* 37, no. 9: 1888–1895. https://doi.org/10.1519/JSC.0000000000004410.

Shalfawi, S. A. I., E. Enoksen, E. Tønnessen and J. Ingebrigtsen. 2014. "Assessing Test-Retest Reliability of the Portable Brower Speed Trap II Testing System." *Kinesiology* 44: 24–30. https://www.researchgate.net/publication/235761587_Assessing_test-retest_reliability_of_the_portable_Brower_Speed_Trap_II_testing_system.

Silva, R., M. Rico-González, R. Lima, Z. Akyildiz, J. Pino-Ortega and F. M. Clemente. 2021. "Validity and Reliability of Mobile Applications for Assessing Strength, Power, Velocity and Change-of-Direction: A Systematic Review." *Sensors* 21, no. 8: 2623. https://doi.org/10.3390/S21082623.

Sugiyama, T., S. Maeo, T. Kurihara, H. Kanehisa and T. Isaka. 2021. "Change of Direction Speed Tests in Basketball Players: A Brief Review of Test Varieties and Recent Trends." *Frontiers in Sports and Active Living* 1: 645350. https://doi.org/10.3389/fspor.2021.645350.

Taber, C., C. Butler, V. Dabek, et al. 2023. "Validity and Reliability of a Computer Vision System to Determine Bar Displacement and Velocity." *International Journal of Strength and Conditioning* 3, no. 1. https://doi.org/10.47206/IJSC.V3I1.263.

Taylor, J. B., A. A. Wright, S. L. Dischiavi, M. A. Townsend and A. R. Marmon. 2017. "Activity Demands During Multi-Directional Team Sports: A Systematic Review." *Sports Medicine* 47, no. 12: 2533–2551. https://doi.org/10.1007/s40279-017-0772-5.

Tissera, K., K. A. Shorter, M. Huynh and A. C. Benson. 2024. "Reliability and Validity of the Fulltrack AI Application to Determine Cricket Bowling Line and Length Compared to 3D Motion Capture." *Sports Biomechanics*: 1–17. https://doi.org/10.1080/14763141.2024.2381108.

Turan, M., S. Ulupınar, S. Özbay, C. Gençoğlu, B. Ç. Savaş and İ. İnce. 2022. "Validity and Reliability of 'My Jump App' to Assess Vertical Jump Performance: A Meta-Analytic Review." https://doi.org/10.21203/rs.3.rs-2154724/v1.