Science & Sports 39 (2024) 654—664

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

ORIGINAL ARTICLE

Reliability of time-motion analysis in striking combat sports

Fiabilité de l'analyse temps-mouvement dans les sports de combat avec des frappes

L. Rohner[a,b,*], C.R. Abbiss[a], W. Poon[a], O.R. Barley[a]

[a] *School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA, Australia*
[b] *Institute of Sports and Preventive Medicine, Saarland University, Saarbruecken, Germany*

Reçu le 9 août 2022 ; accepté le 10 décembre 2023
Disponible sur Internet le 9 octobre 2024

KEYWORDS
Reliability ;
Time-motion analysis ;
Technical-tactical analysis ;
Combat sports ;
Significant strikes

Summary
Introduction. — Time-motion analysis (TMA) is a noninvasive performance analysis technique used in various sports to provide broad insights into the technical, tactical, and physiological demands of competitions by quantifying the mode, frequency, and duration of discrete activities.
Objectives. — This study aimed to assess the intra- and inter-observer reliability of activities in full-contact kickboxing athletes determined using TMA.
Equipment and methods. — A total of 29 bouts from the World Association of Kickboxing Organizations World Championships 2019, involving 58 male athletes, were analyzed by two expert analysts in a randomised order to determine the number of actions performed (i.e., punches, kicks, significant punches, significant kicks) and time (in seconds) spent in different activity zones (i.e., high-intensity action, low-intensity action, referee pause). This design was chosen to reflect the common TMA practices in the field. Additionally, one investigator analysed all bouts a second time with a seven-day interval between analyses.
Results. — Good to excellent relative reliability was observed for punches, kicks, and total strikes both within and between raters (intraclass correlation coefficients; ICC between 0.88 and 0.99). Good and poor intra- and inter-observer reliability were found for significant punches, significant kicks, and total significant strikes (ICC: 0.76—0.85 and 0.16—0.44, respectively). Intra- and inter-observer reliability was excellent and moderate to high, respectively, for high-(HIA) and low-intensity action (LIA) (ICC: 0.91—0.93 and 0.65—0.78, respectively).

* Corresponding author at: Institute of Sports and Preventive Medicine, Saarland University, Campus, Bldg. B 8-2, 66123 Saarbruecken, Germany.
Adresse e-mail : luca_rohner@hotmail.com (L. Rohner).

https://doi.org/10.1016/j.scispo.2023.12.004

Conclusion. — TMA in striking combat sports using Hudle SportsCode has been shown to be a reliable method that can provide information of high practical relevance. However, caution should be taken when categorising strikes as 'significant'. Given that the quality of strikes is a key-scoring criterion across a range of combat sports, this outlines serious concerns about the accuracy and reliability of the current judging practices.

MOTS CLÉS
Fiabilité ;
Analyse temps-mouvement ;
Sports de combat ;
Coups significatifs

Résumé

Introduction. — L'analyse temps-mouvement (TMA) est une technique d'analyse de la performance non invasive utilisée dans divers sports pour donner un aperçu des exigences techniques, tactiques et physiologiques des compétitions en quantifiant le mode, la fréquence et la durée d'activités discrètes.

Objectifs. — Cette étude visait à évaluer la fiabilité intra- et inter-observateur des activités des athlètes de kickboxing en plein contact déterminées par TMA.

Matériel et méthodes. — Un total de 29 combats des Championnats du monde World Association of Kickboxing Organizations 2019, impliquant 58 athlètes masculins ont été analysés par deux analystes experts, dans un ordre aléatoire, pour déterminer le nombre d'actions réalisées (c'est-à-dire les coups de poing, les coups de pied, les coups de poing significatifs, les coups de pied significatifs) et le temps (en secondes) passé dans différentes zones d'activité (c'est-à-dire l'action de haute intensité, l'action de faible intensité, la pause de l'arbitre). Cette conception a été choisie pour refléter les pratiques courantes de TMA sur le terrain. De plus, un investigateur a analysé tous les combats une seconde fois avec un intervalle de sept jours entre les analyses.

Résultats. — Une fiabilité relative bonne à excellente a été observée pour les coups de poing, les coups de pied et les coups totaux, à la fois au sein d'un même évaluateur et entre évaluateurs (coefficients de corrélation intraclasse ; ICC entre 0,88 et 0,99). Une bonne et une mauvaise fiabilité intra- et inter-observateurs ont été observées pour les coups de poing significatifs, les coups de pied significatifs et le total des coups significatifs (ICC : 0,76—0,85 et 0,16—0,44, respectivement). La fiabilité intra- et inter-observateur était excellente et modérée à élevée, respectivement, pour les actions de haute (HIA) et de basse intensité (LIA) (ICC : 0,91—0,93 et 0,65—0,78, respectivement).

Conclusion. — TMA dans les sports de combat avec frappe utilisant Hudle SportsCode s'est avérée être une méthode fiable qui peut fournir des informations d'une grande pertinence pratique. Toutefois, il convient d'être prudent lors de la catégorisation des frappes comme « significatives ». Étant donné que la qualité des frappes est un critère de notation essentiel dans toute une série de sports de combat, cela soulève de sérieuses inquiétudes quant à la précision et à la fiabilité des pratiques de jugement actuelles.

1. Introduction

''Combat sports'' is a broad term encompassing a range of sports that involve one-on-one combat. These sports can be divided into grappling, striking and mixed-style disciplines. Grappling disciplines focus on takedowns, chokes, and joint locks without applying striking techniques (e.g., punches and kicks), while striking disciplines may include punches, kicks, knees, and elbows depending on the ruleset. As the name implies, mixed-style disciplines include a mixture of both striking and grappling [1]. Across these sports, the winner is determined by specific criteria depending on the ruleset of the particular discipline [1]. In striking sports, this is typically the individual who either disables his opponent by knock-out (KO) or technical knock-out (TKO) or scores the greatest number of points through successfully executed techniques [2]. One of the most popular striking combat sports is kickboxing. A typical full-contact kickboxing bout consists of three rounds of 2 minutes with a 1-minute rest in between rounds [3]. Fighters are permitted to execute punches and kicks to the torso or head of their opponent [4]. During competition, the frequency of techniques performed, and thus also exercise intensity, obviously varies drastically. It has been shown that the frequency of certain performed offensive actions (e.g., hook punches, roundhouse kicks) and defensive actions (e.g., foot defence, clinch) can discriminate between winners and losers in kickboxing bouts [5,6]. Thus, measures of such technical-tactical ability may help improve competition performance. Although the time spent in high- (HIA) and low-intensity action phases (LIA) has not been shown to discriminate between winners and losers, these parameters may enable the calculation of ratios such as ''effort to pause'' (E:P) and thus still be of high practical relevance in terms of technical-tactical and conditioning preparations [5,7]. However, despite the importance of understanding competition load in combat sports, no clear consensus exists regarding the assessment or classification of intensity.

Time-motion analysis (TMA) is a noninvasive performance analysis technique used in various sports to provide broad insights into the technical, tactical, and physiological demands of competitions by quantifying the mode, frequency, and duration of discrete activities [8]. While several studies have conducted TMA in striking combat sports such as kickboxing [3,5,6,9,10], boxing [11—13], karate [14—16], and taekwondo [8,17], there is a paucity of research examining the inter- and intra-observer reliability of TMA in combat sports. Typically, intra-observer reliability of TMA in kickboxing has shown reasonable reliability (intraclass correlation coefficient [ICC] > 0.80); however, these studies did not examine inter-observer reliability [3,5,6,9]. The only study reporting inter-observer reliability in kickboxing has shown poor reliability (Cohen's Kappa < 0.20) for some of the recorded variables (i.e., total punches, roundhouse kicks, total attacking actions, and total defensive actions), while the rest of the variables tended to be more reliable (Cohen's Kappa: 0.60—0.86) [5]. None of the aforementioned studies thoroughly reported and discussed both intra- and inter-observer reliability measures as well as their respective confidence intervals. Furthermore, within the current literature, it is unclear how time spent in different activity phases such as HIA and LIA was acquired, as it is rarely stated. One possibility is that both were manually measured using TMA. However, a more time-efficient way would be to only measure HIA and to calculate LIA by subtracting HIA and referee pause from the total round duration. The latter approach makes the sum of HIA, LIA and referee pause consistent with the total round duration, whereas the first approach may under- or overestimate the total round time due to measurement inaccuracies. Thus, variables derived from LIA (e.g., E:P ratio), which are of high practical relevance, might differ depending on whether LIA was measured or calculated.

A better understanding of how analysts identify strikes has implications for judging within combat sports. Research has shown that total strikes landed have an influence on the likelihood of winning in combat sports [18—20]. Furthermore, in certain sports, the term ''significant strikes'' is commonly used despite not being clearly defined in all combat sports striking. In professional kickboxing, a clear definition is provided by the World Association of Kickboxing Organizations, which reads ''A punch (or kick) is considered significant when it is executed with power and leads to a visible impact on the opponent's target area above waist (head or body). Punches (or kicks) which are partially deviated or blocked, or simply touches, brushes or pushes an opponent will not be considered as significant'' [4]. However, the assessment is very subjective even if the variables are clearly defined, and therefore, it is unclear how accurate the TMA around significant strikes is. Considering the fact that significant strikes are not even clearly defined in other sports such as Mixed Martial Arts (MMA), this might have large implications on the TMA data collected and the judging effectiveness in striking combat sports. Thus, it is important that a deeper understanding is developed.

Developing a better understanding of TMA in striking combat sports will have many applications within both practical and academic settings, such as optimising the tactical analysis of competition to determine winning strategies more accurately and better understanding the reliability of key measures of output in combat sports research and practice.

Therefore, the aims of the study were to (i) assess the reliability (i.e., intra- and inter-observer) of activity in full-contact kickboxing athletes (i.e., HIA, LIA, referee pause and frequency of actions) measured using time-motion analysis and (ii) determine if there is a difference in the time of LIA if calculated or measured using TMA. Time structure variables such as LIA as well as many techniques used in kickboxing are also used in many other combat sports, and thus, findings from this study will have important implications in other disciplines outside of kickboxing. Kickboxing was chosen as punches, and kicks were allowed. Therefore, this provides not only information on kickboxing itself but potentially also on striking aspects of other sports such as MMA and boxing, allowing this study to have further reaching implications.

2. Methods

2.1. Participants

To assess the intra- and inter-observer reliability of the TMA as well as potential differences between the measured LIA and calculated LIA in striking combat sports, a total of 29 kickboxing fights involving 58 senior male full-contact kickboxers from the World Association of Kickboxing Organizations World Championships 2019 in Antalya (Turkey) were analysed. For the purpose of this study, only bouts from the weight classes 60 to 63.5 kg (n = 8), 63.5—67 kg (n = 7), 67—71 kg (n = 7), and 71—75 kg (n = 7) were considered, as they were the most heavily occupied divisions according to their number of participants at World Championships 2019. Video footage from all fights was taken from online public-access websites (youtube.com; @WAKOKickboxing). Sample size was calculated according to a web-based calculator created by Arifin [21] with minimum acceptable ICC ($\rho 0$), expected ICC ($\rho 1$), significance level (α) and statistical power ($1-\beta$) set as 0.80, 0.90, 0.05, and 80%, respectively.

This study was retrospective, observational and utilised publicly available data (youtube.com; @WAKOKickboxing). The study presented no risk to participants and, as a result, was deemed out of scope for ethical review by the relevant institutional human research ethics committee. Moreover, this research was carried out fully in accordance with the ethical standards of the *International Journal of Exercise Science* [22].

2.2. Protocol

A time-motion system aiming to incorporate the most relevant TMA variables for striking combat sports was elaborated. Each bout consisted of three rounds of 2 minutes with 1 minute of passive rest between rounds. Fighters activity in each round was divided into high-intensity action (HIA), low-intensity action (LIA), and referee pause, according to the approach of Ouergui et al. [3] (Table 1). Moreover, the number of punches, kicks, significant punches, and significant kicks applied by each fighter in each round were recorded

Table 1 Technical-tactical and time-motion variables with the respective criteria.

Variables	Criteria of analysis
Punch	Any punch performed with the intention to hit the opponent. Feints which aim to mislead the opponent by performing a quick movement will not be counted as punch
Kick	Any kick performed with the intention to hit the opponent. Feints which aim to mislead the opponent by performing a quick movement will not be counted as kick
Significant punch	A punch is considered significant when it is executed with power and leads to a visible impact on the opponent's target area above waist (head or body). Punches which are partially deviated or blocked, or simply touches, brushes or pushes an opponent will not be considered as significant [4]
Significant kick	A kick is considered significant when it is executed with power and leads to a visible impact on the opponent's target area above waist (head or body). Kicks which are partially deviated or blocked, or simply touches, brushes or pushes an opponent will not be considered as significant [4]
High-intensity action (HIA)	Ouergui et al. [3] (p. 3538) defined HIA as ''the phase when fast movements were performed to gain a specific position, to defend a favourable position by the opponent, or in an attempt to strike the opponent with power''. HIA phase started when ''1 foot or hand moved to initiate the attack'' and ended when ''(i) the foot/hand that delivered the last kick/punch of the action returned to the floor/guard, (ii) the punching or blocking limb was retracted, (iii) a 'knockdown' followed the technique, or (iv) the referee used the stop hand signal''
Low-intensity action (LIA)	Ouergui et al. [3] (p. 3538) considered LIA as ''stable positions or movements (e.g., foot work: step, slide, bounce; deceptive movements: foot, head/body, or hand feint) requiring little effort to maintain low-speed displacements or movements without opposition''
Referee pause	Ouergui et al. [3] (p. 3538) considered referee pause ''when stoppage time was initiated by the referee's 'break' hand signal or if a competitor was knocked to the ground. Stoppage time ceased on the referee's hand signal to 'begin'''

according to the definitions of the World Association of Kickboxing Organizations ruleset (Table 1). Additionally, the total duration of each round was recorded, and total strikes (sum of punches and kicks) as well as total significant strikes (sum of significant punches and kicks) were calculated.

Two expert analysts with greater than six years of competitive experience in striking combat sports analysed all videos to determine the number of actions performed (i.e., punches, kicks, significant punches, significant kicks) and the time (in seconds) spent in different activity zones (i.e., HIA, LIA, referee pause). This design is comparable to most TMA studies in combat sports currently using one to two analysers for their data collection [3,6—9,12,23—25]. Video footage of each round and each fighter was analysed individually using the TMA system scripted with the software Hudle SportsCode (Agile Sports Technologies Inc, Lincoln, Nebraska, USA). All fights were randomised but analysed in the same randomised order by each investigator. One investigator (rater 1) analysed all matches a second time with a seven-day interval between analyses. Investigators viewed the footage generally at one-quarter or half of normal speed. However, they were permitted to rewind the footage and watch events frame-by-frame to ensure accuracy.

To become familiar with the system, both investigators went through a familiarisation phase before starting the analysis of the actual matches. Three matches, from the same championship as the ones utilised in the analysis, were analysed during this phase. The TMA results of both investigators (i.e., time spent in different activity zones and number of actions performed) were compared, and some specific match sequences were discussed to ensure that both agreed on the definitions of different activities. Furthermore, both investigators determined and standardised the exact order of analysis for the different variables of the TMA system together (i.e., 1. Total duration, 2. Referee pause, 3. HIA red corner, etc.).

2.3. Statistical analysis

Descriptive statistics are presented as the means ± standard deviations (SD), while inference statistics are accompanied by 95% confidence intervals (CI). The intraclass correlation coefficient (ICC) was reported as a measure of relative reliability. Intra-observer reliability estimates were calculated based on a single measurement, absolute-agreement, two-way mixed effects model (3,1). For inter-observer reliability estimates, a single rater, absolute-agreement, two-way random effects model (2,1) was used [26]. Calculations were carried out using IBM SPSS Statistics (Version 26) predictive analytics software. ICCs of < 0.50, 0.50—0.75, 0.75—0.90, and > 0.90 were considered poor, moderate, good, and excellent reliability, respectively [26]. Furthermore, the typical error (TE) of measurement in raw units, expressed as the coefficient of variation (CV%), was calculated as an indicator of absolute reliability. CV% were determined using a spreadsheet provided by Hopkins [27]. For the reliability

Table 2 Measures of intra-observer reliability of technical-tactical and time-motion variables per round.

Variables	Observation 1 mean (±SD)	Observation 2 mean (±SD)	Change in mean [95% CI]	TE [95% CI]	CV% [95% CI]	ICC [95% CI]
Punches	30.0 (±10.4)	29.7 (±10.3)	−0.3 [−0.7, 0.2]	2.0 [1.8, 2.2]	6.6 [5.9, 7.3]	0.96 [0.95, 0.97]
Kicks	12.4 (±5.5)	12.4 (±5.3)	−0.1 [−0.1, 0.0]	0.5 [0.4, 0.5]	3.7 [3.3, 4.1]	0.99 [0.99, 0.99]
Total strikes	42.4 (±11.4)	42.1 (±11.4)	−0.3 [−0.7, 0.1]	2.0 [1.8, 2.2]	4.8 [4.3, 5.3]	0.97 [0.96, 0.98]
Sign. punches	3.1 (±2.6)	3.5 (±2.8)	0.4 [0.2, 0.6]	1.0 [0.9, 1.1]	30.4 [27.5, 34.0]	0.85 [0.80, 0.89]
Sign. kicks	1.4 (±1.3)	1.6 (±1.4)	0.2 [0.1, 0.3]	0.6 [0.6, 0.7]	43.6 [39.4, 48.7]	0.76 [0.69, 0.82]
Total sign. strikes	4.5 (±3.0)	5.1 (±3.1)	0.6 [0.3, 0.8]	1.2 [1.1, 1.4]	25.9 [23.4, 29.0]	0.82 [0.75, 0.87]
HIA (s)	37.6 (±9.5)	35.1 (±8.3)	−2.5 [−3.0, −2.1]	2.0 [1.8, 2.2]	5.5 [5.0, 6.2]	0.91 [0.62, 0.97]
LIA (s)	84.4 (±10.8)	86.9 (±9.8)	2.5 [2.1, 2.9]	2.0 [1.8, 2.3]	2.4 [2.1, 2.7]	0.93 [0.70, 0.97]
Referee pause (s)	9.1 (±17.0)	9.1 (±17.0)	0.0 [−0.1, 0.0]	0.1 [0.1, 0.1]	1.4 [1.3, 1.6]	1.00 [1.00, 1.00]
Total duration (s)	132.3 (±15.6)	132.3 (±15.5)	0.0 [0.0, 0.0]	0.1 [0.1, 0.1]	0.1 [0.1, 0.1]	1.00 [1.00, 1.00]

SD: standard deviation; CI: confidence intervals; TE: typical error; CV: coefficient of variation; ICC: intraclass correlation coefficient; HIA: high-intensity action; LIA: low-intensity action.

assessment, the values of each variable for each individual round were used to minimize cumulative errors, as overall errors may appear less incorrect than they actually are [28]. For the assessment of inter-observer reliability, only the initial analysis of rater 1 was taken into account and compared to the analysis of rater 2. Additionally, intra- and inter-observer agreement for the most relevant variables, namely, total strikes, total significant strikes, HIA, and LIA, are illustrated as Bland—Altman plots with 95% limits of agreement (LoA). Bland—Altman plots allow visual inspection of the data with regard to systematic bias and random error. A linear mixed effects model (LMM) was run to determine whether there was a difference between the two types of LIA (measured and calculated) and how they interacted with the variable rater (rater 1 and rater 2) after controlling for variation in random effects. The model was fitted using time (LIA) as the dependent variable and type and rater as well as their interaction as fixed effects. Rounds nested within subjects and replicates nested within raters were included as random effects. The formula of the model was as follows:

$$\text{Time} \sim \text{Type} * \text{Rater} + \left(1 \,\middle|\, \text{SubjectID}/\text{Round}\right) + \left(1 \,\middle|\, \text{Rater}/\text{Replicate}\right)$$

The LMM allowed us to include both analyses of rater 1 as well as the analysis of rater 2; therefore, no data were lost. Bland—Altman plots and the LMM were computed using the R statistics program (R Core Team 2020). Alpha was set at $P \leq 0.05$.

3. Results

3.1. Reliability

All variables except for the significant strike variables showed excellent intra-observer reliability (ICC $\geq$ 0.91; Table 2). Significant strike variables (i.e., significant punches, kicks and total strikes) revealed good ICCs but high CVs (ICC: 0.76—0.85; CV: > 25%; Table 2).

Strike variables (i.e., punches, kicks, total strikes) showed good to excellent inter-observer reliability (ICC $\geq$ 0.88; Table 3). Significant strike variables revealed poor inter-observer reliability (ICC $\leq$ 0.44; Table 3). Time structure variables showed good to excellent inter-observer reliability (ICC $\geq$ 0.78; Table 3), except LIA, which showed moderate reliability (ICC = 0.65; Table 3).

Bland—Altman plots, created for the variables total strikes, total significant strikes, HIA, and LIA, revealed the extent of within- and between-rater systematic biases as well as random errors (Figs. 1—4).

3.2. Measured LIA vs. calculated LIA

The measured LIA was significantly lower ($\beta = -1.26$, $P < 0.001$) than the calculated LIA, taking the values of both rater 1 (86.9 ± 10.0 s and 85.6 ± 10.4 s, respectively) and rater 2 (86.0 ± 10.5 s and 76.7 ± 11.2 s, respectively) together. Additionally, an interaction effect was observed between type (measured or calculated) and rater ($\beta = -8.03$, $P < 0.001$).

Table 3 Measures of inter-observer reliability of technical-tactical and time-motion variables per round.

Variables	Rater 1 mean (±SD)	Rater 2 mean (±SD)	Change in mean [95% CI]	TE [95% CI]	CV% [95% CI]	ICC [95% CI]
Punches	30.0 (±10.4)	28.0 (±9.4)	−2.0 [−2.7, −1.3]	3.2 [2.9, 3.6]	11.0 [10.0, 12.3]	0.88 [0.80, 0.92]
Kicks	12.4 (±5.5)	12.5 (±5.5)	0.1 [−0.1, 0.3]	1.1 [1.0, 1.2]	8.8 [7.9, 9.8]	0.96 [0.95, 0.97]
Total strikes	42.4 (±11.4)	40.5 (±10.6)	−1.9 [−2.6, −1.1]	3.4 [3.1, 3.8]	8.2 [7.4, 9.2]	0.89 [0.83, 0.93]
Sign. punches	3.1 (±2.6)	1.9 (±1.7)	−1.2 [−1.5, −0.9]	1.6 [1.4, 1.7]	61.5 [55.6, 68.7]	0.44 [0.22, 0.60]
Sign. kicks	1.4 (±1.3)	0.4 (±0.7)	−1.0 [−1.2, −0.8]	0.9 [0.8, 1.0]	101.3 [91.7, 113.3]	0.16 [−0.02, 0.32]
Total sign. strikes	4.5 (±3.0)	2.3 (±1.9)	−2.2 [−2.6, −1.8]	1.9 [1.7, 2.1]	55.0 [49.8, 61.5]	0.31 [0.02, 0.53]
HIA (s)	37.6 (±9.5)	37.7 (±9.4)	0.1 [−0.8, 1.1]	4.5 [4.0, 5.0]	11.8 [10.7, 13.2]	0.78 [0.71, 0.83]
LIA (s)	84.4 (±10.8)	76.7 (±11.3)	−7.7 [−8.7, −6.7]	4.8 [4.4, 5.4]	6.0 [5.4, 6.7]	0.65 [0.03, 0.85]
Referee pause (s)	9.1 (±17.0)	8.9 (±16.8)	−0.2 [−0.3, 0.0]	0.7 [0.6, 0.8]	7.7 [7.0, 8.6]	1.00 [1.00, 1.00]
Total duration (s)	132.3 (±15.6)	132.5 (±15.5)	0.2 [0.2, 0.3]	0.3 [0.3, 0.4]	0.3 [0.2, 0.3]	1.00 [1.00, 1.00]

SD: standard deviation; CI: confidence intervals; TE: typical error; CV: coefficient of variation; ICC: intraclass correlation coefficient; HIA: high-intensity action; LIA: low-intensity action.

Figure 1 Bland—Altman plots of the variable total strikes illustrating (a) the agreement within rater 1 (between observation 1 and 2) and (b) between rater 1 (observation 1) and rater 2. Individual differences are plotted against their respective means. The bias line and random error lines forming the 95% limits of agreement (+1.96 SD, −1.96 SD) are presented as dashed lines, while their respective 95% confidence intervals are characterised by dotted lines. The red line represents the zero-difference line.

4. Discussion

The present study investigated the intra- and inter-observer reliability of activities in full-contact kickboxing athletes recorded using TMA. A second aim of the study was to determine if there is a difference between the manually measured LIA and the LIA when it was calculated by subtracting the referee pause and the HIA from the total duration of the par-

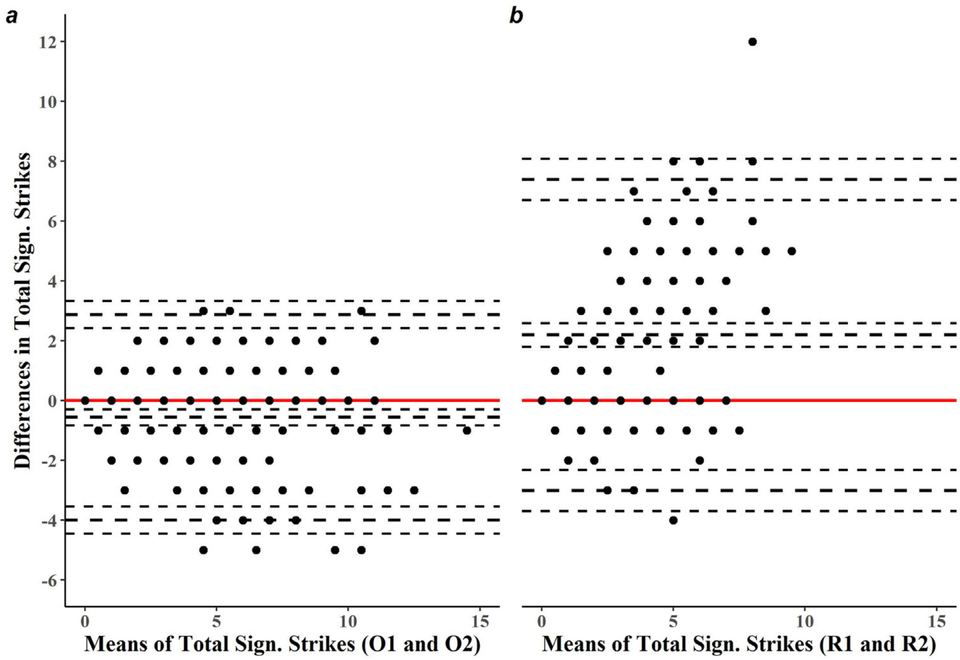

Figure 2 Bland—Altman plots of the variable total significant strikes illustrating (a) the agreement within rater 1 (between observations 1 and 2) and (b) between rater 1 (observation 1) and rater 2. Individual differences are plotted against their respective means. The bias line and random error lines forming the 95% limits of agreement (+1.96 SD, −1.96 SD) are presented as dashed lines, while their respective 95% confidence intervals are characterised by dotted lines. The red line represents the zero-difference line.

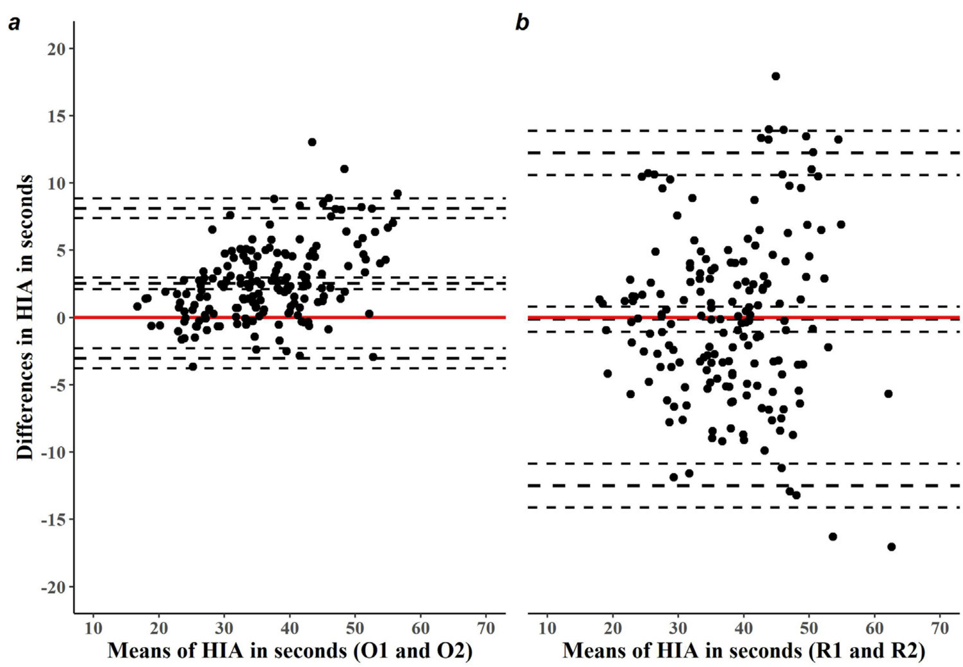

Figure 3 Bland—Altman plots of the variable high-intensity action (HIA) illustrating (a) the agreement within rater 1 (between observations 1 and 2) and (b) between rater 1 (observation 1) and rater 2. Individual differences are plotted against their respective means. The bias line and random error lines forming the 95% limits of agreement (+1.96 SD, −1.96 SD) are presented as dashed lines, while their respective 95% confidence intervals are characterised by dotted lines. The red line represents the zero-difference line.

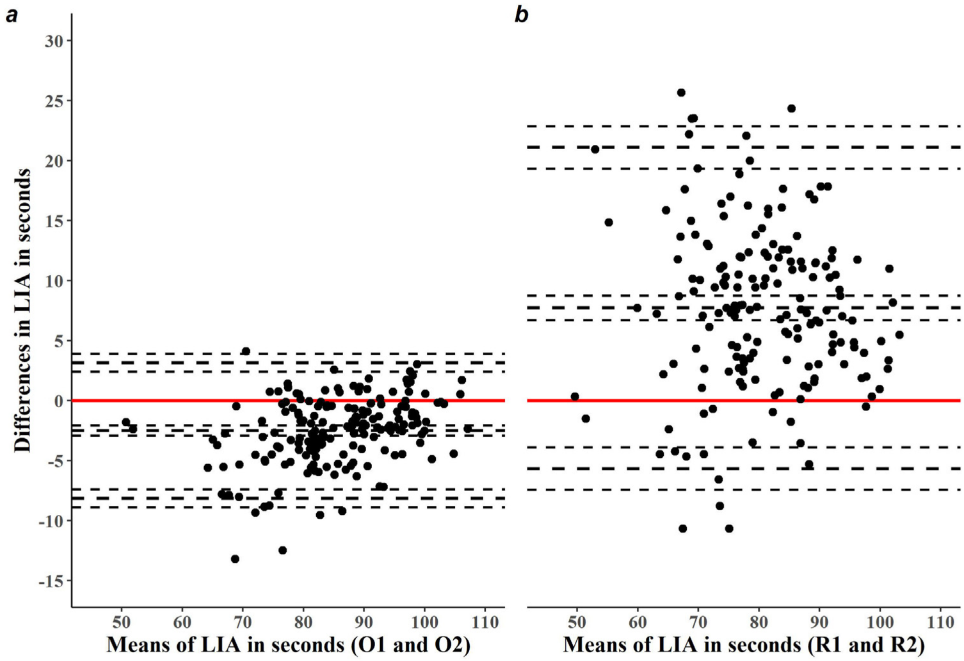

Figure 4 Bland—Altman plots of the variable low-intensity action (LIA) illustrating (a) the agreement within rater 1 (between observations 1 and 2) and (b) between rater 1 (observation 1) and rater 2. Individual differences are plotted against their respective means. The bias line and random error lines forming the 95% limits of agreement (+1.96 SD, −1.96 SD) are presented as dashed lines, while their respective 95% confidence intervals are characterised by dotted lines. The red line represents the zero-difference line.

ticular round. Although previous studies have examined TMA in striking combat sports, a novel aspect of this study was the thorough assessment of both intra- and inter-observer reliability of TMA variables reporting absolute and relative reliability coefficients as well as the reliability of the frequently used ''significant strikes'' variable. Moreover, this study also quantified all variables per round and, thus, can also be considered a performance analysis study. The main findings are that (i) frequency of actions (i.e., punches, kicks, total strikes) were found to be reliable both within and between raters (ICC ≥ 0.88; Tables 2 and 3); (ii) frequency of significant actions (i.e., significant punches, kicks, total strikes) showed good intra- (ICC: 0.76—0.85; Table 2) and poor inter-observer reliability (ICC ≤ 0.44; Table 3); (iii) time spent in different activity zones (i.e., HIA, LIA, referee pause) and total round duration were generally found to be reliable both within and between raters (ICC ≥ 0.78; Tables 2 and 3), except the variable LIA showing moderate inter-observer reliability (ICC = 0.65); and (iv) measured LIA was significantly lower compared with calculated LIA ($\beta = -1.26$, $P < 0.001$), indicating that there is a tendency of either LIA, HIA or both being underestimated when they are measured. Additionally, an interaction effect between type (measured or calculated) and rater ($\beta = -8.03$, $P < 0.001$) was found, meaning that the extent of the difference between measured LIA and calculated LIA depends on the rater.

Accuracy and intra- and inter-observer reliability are of high importance in applied performance analysis, as the data are used by coaches to provide feedback and guide athlete performance [29]. The data may also have important implications for judging, which requires the identification of such actions (i.e. total strikes landed and the quality of said strikes) for scoring purposes. Within the present study, high intra-observer reliability for frequencies of actions (i.e., punches, kicks, total strikes) was observed (Table 2), which is comparable to that observed in previous research investigating kickboxing bouts (ICC ≥ 0.90) [5,6]. The good (except in one case, moderate) inter-observer reliability for frequencies of actions found in this study (Table 3) is comparable to the inter-observer agreement for frequencies of actions previously observed in taekwondo (Cohen's Kappa = 0.95) [8]. However, previous authors have also shown poor inter-observer reliability for similar variables in kickboxing bouts (Cohen's Kappa: 0—0.20) [5]. The focus of the aforementioned study was clearly not on reliability. The results of the reliability analyses were only mentioned in the methods section, and, therefore, there was no further discussion as to why the values were so low. The Bland—Altman plot of the present study showed a general trend for rater 1 to count more strikes than rater 2 (+1.9; Fig. 1). Due to the dynamic nature of many sports, misclassifications are a reappearing issue in performance analysis despite clear definitions of different actions [30,31]. Nevertheless, TMA seems to be a reliable method for determining the frequencies of actions in striking combat sports.

In the present study, significant strike variables (i.e., significant punches, kicks, total strikes) revealed a significant drop in reliability when compared to frequencies of actions, particularly between raters. The approach used in the present study for assessing the reliability of significant strikes in combat sports over a sample size of 29 bouts is novel. We are only aware of one other study in which the reliability of strike outcomes (successful, partially successful, unsuccessful) in combat sports was assessed [31].

However, in this study an approach proposed by Cooper et al. [32] which involved dividing a single boxing bout into 36 × 10-second time cells and using frequency counts of each time cell for comparisons within and between raters has been used. Reliability was assessed using the proportion of perfect agreement (PA). Intra- and inter-observer PA for strike outcome ranged between 75 and 100% and was even 100% when considering PA ± 1, indicating good reliability. It is difficult to compare these data with those of the present study, as different approaches and measurements of reliability were used. Considering the importance of significant strikes in terms of a match outcome, our results indicating poor inter-observer reliability are problematic. Although the definitions of significant strikes from the official World Association of Kickboxing Organizations ruleset were used, investigators rarely counted the exact same number of significant strikes (Fig. 2b). This clearly shows how difficult it is to decide whether a strike is significant or not purely based on subjective criteria and without any objective measurement such as strike power or impact on the opponent's body. These results are important not only for determining strike variables within performance analysis but also for determining competition outcomes where decisions are based on subjective decisions by judges. In fact, inconsistencies between judges in scoring Muay Thai bouts have previously been reported [33]. Hahn et al. [34] discussed the problems of the subjective scoring process in amateur boxing and believed that an automated scoring system should be considered to avoid these problems. In addition to combat sports, many other sports (e.g., surfing, gymnastics, freestyle skiing) are facing similar difficulties. Indeed, performance within approximately one-third of all sports federations recognised by the International Olympic Committee (IOC) is to some extent determined by a subjective judging system [35]. It has been shown, for instance, that judges assessing the five-gymnast ensemble in an ideal environment were only able to detect 20—40% of all ''true errors'' (determined based on a consensus of the five most experienced international judges), which is far below optimum [36]. Bearing in mind the abovementioned considerations, subjective variables such as significant strikes within combat sports analysis should be interpreted with caution.

Time structure variables whose beginning and ending were usually indicated by a clear hand signal (i.e., total duration and referee pause) showed, as expected, almost perfect intra- and inter-observer reliability (Tables 2 and 3). Both HIA and LIA revealed high intra-observer reliability (Table 2), agreeing with previous TMA literature in combat sports (ICCs > 0.80) [3,5,7,9,14,24]. Only two of these studies [7,24] have also examined inter-observer reliability, with very high reliability reported (ICCs > 0.98). However, the HIA and LIA in the present study revealed only moderate to good inter-observer reliability. While it is unclear which analysis software was used by Miarka et al. [24], Del Vecchio et al. [10] analysed their matches using a DVD player and a chronometer. Actually, one would rather expect higher reliability from an analysis software such as Hudle SportsCode than from a DVD player and a chronometer. The former has specifically been created for such analyses and allows sequences to be watched in slow motion and rewound easily to exactly determine the beginning/ending of activity phases. Bland—Altman plots within the present study indicate that the more intensive a round was, the greater the differences tended to be both within and between raters for HIA and LIA (Figs. 3 and 4). Furthermore, Bland—Altman plots did not reveal any systematic error between raters 1 and 2 for HIA; however, there was an average difference of 7.7 seconds between raters for LIA. As a result, the LIA was only moderately reliable compared to the good relative reliability of the HIA. Nevertheless, the absolute reliability of the LIA was higher than that of the HIA. This can be explained by the fact that the CV is calculated based on the variability of the differences between raters 1 and 2 and does not take into account systematic errors [37]. One limitation of the present study is that there was only one camera perspective available to analyse the videos, and therefore, the view was sometimes obstructed by either one of the fighters or by the referee. This potentially could have affected some analyses. However, previous studies using TMA in combat sports also used only one camera perspective [3,6—9,12,23—25]. Moreover, during competition in most combat sports, referees have to assess everything in real time, not even allowing them to consider a slow motion and replay. There is a clear need for future research to determine the best practices for TMA in combat sports and the reliability of different methodologies, which has been surprisingly neglected given the frequent use of TMA in combat sports. Despite the abovementioned issues, time structure variables such as HIA and LIA are valuable, especially considering their high relevance for practitioners. They can use these data to obtain a better understanding of the requirements of a certain combat sport and to plan training sessions more specifically.

There is a small but significant difference between the measured LIA and calculated LIA ($\beta = -1.26$, $P < 0.001$). Moreover, the difference in time between calculated and measured values varies depending on the rater. We are unaware of any previous research that has investigated the difference between the measured LIA and calculated LIA. Additionally, it is not typically reported within studies if HIA and LIA are directly measured or if LIA is calculated by subtracting HIA and referee pause from the total duration of the round. Taking these results together with the reliability measures, it seems as if it would be appropriate to measure HIA and calculate LIA since HIA, referee pause, and total round duration showed better inter-observer ICCs and less systematic bias than manually measured LIA.

5. Conclusion

In conclusion, frequencies of action (i.e., punches, kicks, total strikes) were generally found to be reliable both within and between raters. However, caution should be taken when categorising strikes as 'significant', as the present study found such variables (i.e., significant punches, kicks, total strikes) to have only poor inter-observer reliability. Given that the quality of strikes is a key-scoring criterion across a range of combat sports, this outlines serious concerns about the accuracy and reliability of the current judging practices. Furthermore, time structure variables (i.e., HIA, LIA, referee pause, total duration) were found to be reliable both within and between raters, except LIA, which showed an ICC corresponding only to moderate inter-observer reliability. This might have been caused by the fairly high systematic bias between raters 1 and 2. The measured LIA was significantly lower than the calculated LIA. Taking this result together with the reliability results, it can be concluded that the calculated LIA is the more appropriate and at the same time the more time-efficient way to acquire the LIA. Apart from significant strike variables, TMA in striking combat sports using Hudle SportsCode has been shown to be a reliable method that can provide information of high practical relevance. A potential limitation of the present study is the exclusion of bouts from the lightest (−51, −54, −57, −60 kg) and heaviest (−81, −86, −91, +91 kg) weight classes, limiting the analysis to the middle range of weight classes. In the lower weight classes, fights could plausibly be faster and have more actions, while the opposite may be true for the higher weight classes. Since the investigators viewed the footage generally at one-quarter or half of normal speed, the authors maintain the view that this would not have affected the results. The present study focused on the middleweight classes as they were the most heavily occupied divisions according to their number of participants at World Championships 2019. Future research is needed to confirm the results of the present study and to investigate reliability, especially in combat sports such as Muay Thai and MMA, since clinching, and in the case of MMA also grappling, might have an influence on the reliability of different TMA variables.

Disclosure of interest

The authors declare that they have no competing interest.

Références

[1] Barley OR, Chapman DW, Abbiss CR. The current state of weight-cutting in combat sports-weight-cutting in combat sports. Sports (Basel) 2019;7(5).

[2] Barley OR, Chapman DW, Guppy SN, Abbiss CR. Considerations when assessing endurance in combat sport athletes. Front Physiol 2019;10:205.

[3] Ouergui I, Hssin N, Haddad M, Franchini E, Behm DG, Wong DP, et al. Time-motion analysis of elite male kickboxing competition. J Strength Cond Res 2014;28(12):3537—43.

[4] WAKO Full contact Rules [Internet]. WAKO Kickboxing; 2016.

[5] Ouergui I, Davis P, Houcine N, Marzouki H, Zaouali M, Franchini E, et al. Hormonal, physiological, and physical performance during simulated kickboxing combat: differences between winners and losers. Int J Sport Physiol Perform 2016;11(4):425—31.

[6] Ouergui I, Hssin N, Franchini E, Gmada N, Bouhlel E. Technical and tactical analysis of high level kickboxing matches. Int J Perform Anal Sport 2013;13(2):294—309.

[7] Del Vecchio FB, Hirata SM, Franchini E. A review of time-motion analysis and combat development in mixed martial arts matches at regional level tournaments. Percept Mot Skills 2011;112(2):639—48.

[8] Bridge CA, Jones MA, Drust B. The activity profile in international Taekwondo competition is modulated by weight category. Int J Sports Physiol Perform 2011;6(3):344—57.

[9] Ouergui I, Benyoussef A, Houcine N, Abedelmalek S, Franchini E, Gmada N, et al. Physiological responses and time-motion analysis of kickboxing: differences between full contact, light contact, and point fighting contests. J Strength Cond Res 2019;35(9):2558—63.

[10] Silva JJR, Del Vecchio FB, Picanço LM, Takito MY, Franchini E. Time-motion analysis in Muay-Thai and Kick-Boxing amateur matches. J Hum Sport Exerc 2011;6(3):490—6.

[11] Davis P, Benson PR, Pitty JD, Connorton AJ, Waldock R. The activity profile of elite male amateur boxing. Int J Sports Physiol Perform 2015;10(1):53—7.

[12] Davis P, Wittekind A, Beneke R. Amateur boxing: activity profile of winners and losers. Int J Sports Physiol Perform 2013;8(1):84—91.

[13] Dunn EC, Humberstone CE, Iredalel KF, Martine DT, Blazevich AJ. Human behaviours associated with dominance in elite amateur boxing bouts: a comparison of winners and losers under the Ten Point Must System. PLoS One 2017;12(12):84—92.

[14] Chaabène H, Franchini E, Miarka B, Selmi MA, Mkaouer B, Chamari K. Time—motion analysis and physiological responses to karate official combat sessions: is there a difference between winners and defeated karatekas? Int J Sports Physiol Perform 2014;9(2):302—8.

[15] Tabben M, Coquart J, Chaabène H, Franchini E, Ghoul N, Tourny C. Time-motion, tactical and technical analysis in top-level karatekas according to gender, match outcome and weight categories. J Sports Sci 2015;33(8):841—9.

[16] Tabben M, Miarka B, Chamari K, Beneke R. Decisive moment: a metric to determine success in elite karate bouts. Int J Sports Physiol Perform 2018;13(8):1000—4.

[17] Tornello F, Capranica L, Chiodo S, Minganti C, Tessitore A. Time-motion analysis of youth Olympic Taekwondo combats. J Strength Cond Res 2013;27(1):223—8.

[18] Ashker SE. Technical and tactical aspects that differentiate winning and losing performances in boxing. Int J Perform Anal Sport 2017;11:356—64.

[19] Collier T, Johnson AL, Ruggiero J. Aggression in mixed martial arts: an analysis of the likelihood of winning a decision [Internet]. In: Jewell RT, editor. Violence and aggression in sporting contests: economics, history and policy. New York, NY: Springer; 2012 [cited 2022 Mar 11].

[20] James LP, Robertson S, Haff GG, Beckman EM, Kelly VG. Identifying the performance characteristics of a winning outcome in elite mixed martial arts competition. J Sci Med Sport 2017;20(3):296—301.

[21] Arifin WN. A web-based sample size calculator for reliability studies. 2018;10:67—76.

[22] Navalta J, Stone W, Lyons T. Ethical issues relating to scientific discovery in exercise science. Int J Exerc Sci 2019;12(1):1—8.

[23] James LP, Sweeting AJ, Kelly VG, Robertson S. Longitudinal analysis of tactical strategy in the men's division of the ultimate fighting championship. Front Artif Intell 2019;2.

[24] Miarka B, Coswig VS, Vecchio FBD, Brito CJ, Amtmann J. Comparisons of time-motion analysis of mixed martial arts rounds by weight divisions. Int J Perform Anal Sport 2015;15(3):1189—201.

[25] Ouergui I, Houcine N, Marzouki H, Davis P, Franchini E, Gmada N, et al. Physiological responses and time-motion analysis of small combat games in kickboxing: impact of ring size and number of within-round sparring partners. J Strength Cond Res 2017;31(7):1840—6.

[26] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155—63.

[27] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30(1):1—15.

[28] Hughes M. The essentials of performance analysis: an introduction. 1st ed. New York: Routledge; 2007.

[29] Hughes M, Franks IM, Dancs H, editors. Essentials of performance analysis in sport: third edition. 3rd ed. New York: Routledge; 2019.

[30] Hughes M, Cooper S-M, Nevill A, Brown S. An example of reliability testing and establishing performance profiles for

non-parametric data from performance analysis. Int J Comput Sci Sport 2003;2:34—56.
[31] Thomson E, Lamb K, Nicholas C. The development of a reliable amateur boxing performance analysis template. J Sports Sci 2013;31(5):516—28.
[32] Cooper S-M, Hughes M, O'Donoghue P, Nevill MA. A simple statistical method for assessing the reliability of data entered into sport performance analysis systems. Int J Perform Anal Sport 2007;7(1):87—109.
[33] Myers T, Nevill A, Al-Nakeeb Y. An examination of judging consistency in a combat sport. J Quan Anal Sports 2010;6:3.
[34] Hahn AG, Helmer RJN, Kelly T, Partridge K, Krajewski A, Blanchonette I, et al. Development of an automated scoring system for amateur boxing. Proc Eng 2010;2(2):3095—101.
[35] Stefani R. Predicting outcomes. In: Bennett J, editor. Statistics in sport. New York: Hodder Education Publishers; 1998.
[36] Flessas K, Mylonas D, Panagiotaropoulou G, Tsopani D, Korda A, Siettos C, et al. Judging the judges' performance in rhythmic gymnastics. Med Sci Sports Exerc 2015;47(3):640—8.
[37] Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: F.A. Davis; 2015.