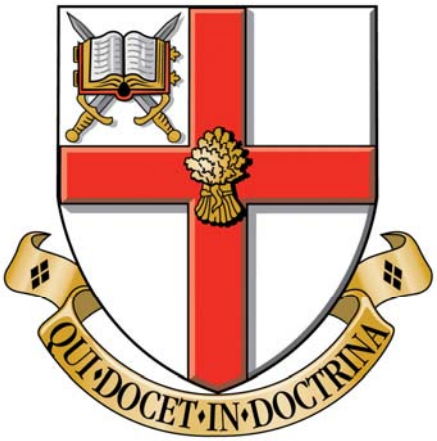

University of Chester

Author(s): Edward Thomson ; Kevin Lamb ; Ceri Nicholas

Title: The development of a reliable amateur boxing performance analysis template

Date: 2013

Originally published in: Journal of Sports Sciences

Example citation: Thomson, E., Lamb, K., & Nicholas, C. (2013). The development of a reliable amateur boxing performance analysis template. *Journal of Sports Sciences, 31*(5), 516-528. doi: http://dx.doi.org/10.1080/02640414.2012.738922

Version of item: Authors' accepted manuscript

Available at: http://hdl.handle.net/10034/335797

The Development of a Reliable Amateur Boxing Performance Analysis Template.

Keywords: performance analysis, reliability, amateur boxing.

Abstract

The aim of this study was to devise a valid performance analysis system for the assessment of the movement characteristics associated with competitive amateur boxing and assess its reliability. Key performance indicators to characterise the demands of an amateur contest (offensive, defensive and feinting) were developed and notated using a computerized notational analysis system. Data was subjected to intra- and inter-observer reliability assessment via the methods outlined by Cooper et al. (2007), which focus on the frequencies of events between observations. For all performance indicators, intra-observer reliability revealed non-significant differences between observations ($P > 0.05$) and high agreement was established (80 - 100%) regardless of whether *exact* or the *reference value* of ±1 was applied. Inter-observer reliability was less impressive for both analysts (amateur boxer & experienced analyst), with the proportion of agreement ranging from 33–100%. Nonetheless, there was no systematic bias between observations for any indicator ($P > 0.05$), and the proportion of agreement within the reference range (±1) was 100%. A reliable performance analysis template has been developed for the assessment of amateur boxing performance and is available for use by researchers, coaches and athletes to classify and quantify the sport-specific demands.

1. Introduction

Performance analysis is concerned with classifying and quantifying the essential aspects of sports performance and potentially can provide detailed feedback to competitors and coaches for the purpose of improving future performances (Hughes & Bartlett, 2002; Hughes & Franks, 2004). In the past ten years it has been applied to a variety of sports for assessing the technical, tactical and physical aspects of players' performances (Hughes, 2004; Hughes & Franks, 2004; Carling, Williams & Reilly, 2005). To date, team sports have received the most attention from performance analysts, in particular soccer (Bloomfield, Polman & O'Donoghue, 2007; Tenga, Kanstad, Ronglan & Bahr, 2009; Clark, 2010), rugby (Sykes, Twist, Hall, Nicholas & Lamb, 2009; Vaz, Mouchet, Carreras & Morente, 2011) and volleyball (Hughes & Daniel, 2003; Drikos & Vagenas, 2011), though individual sports, such as racket sports (O'Donoghue & Ingram, 2001; O'Donoghue, 2002), athletics (Brown, 2005; Brown & O'Donoghue, 2007) and combat sports (Atan & Imamoglu, 2005; Nunan, 2006; Laird & McLeod, 2009), have also received scrutiny. A combat sport that has received little attention is boxing, which is surprising given the observable demands of its dynamic, competitive environment and the pivotal emphasis on tactical manoeuvres. Moreover, boxing is popular at professional and amateur levels; in the United Kingdom alone, nearly 140,000 individuals participate in amateur boxing at least once per week (Sport England, 2008) and approximately 21,500 of them are registered for competitive boxing ((Amateur Boxing Association of England) ABAE, 2009). It is popular world-wide, with nearly 200 nations affiliated to its international governing body, the Amateur Boxing International Association (AIBA) (Smith & Draper, 2007).

Recent changes to the scoring mechanism in amateur boxing (ABAE, 2009; AIBA, 2008) mean that competitors now are rewarded an unlimited amount of points for landing blows (hits) of 'sufficient' force upon the opponent target area, whereas previously scoring (and the outcome of a contest) was based on impressionistic judgements (Partridge, Hayes, James, Hill, Gin & Hahn, 2005; Smith,

2006; AIBA, 2008). This has subsequently led to alterations in the tactics of boxers within a contest and placed a greater emphasis on landing single, forceful blows from smaller combinations of punches (Smith, 2006) rather than throwing combinations containing many punches. In addition, the work-to-rest ratios for elite and 'open' class amateur boxers have been altered from 4 x 2 minute rounds to 3 x 3 minute rounds (AIBA, 2008), which is likely to have had an impact on the boxers' activity patterns within rounds and the accompanying physiological responses. Given the emerging pre-eminence of performance analysis, these variations to the sport provide an enticing opportunity for the development of a boxing-specific model that will inform coaches and their fighters in the manner established in other sports. Nonetheless, regardless of the sophistication of such an analysis, the data generated needs to be valid and reliable. That is, the observation and subsequent classification of the characteristics of the sport (its 'performance indicators') need to be comprehensive and the act of recording such events needs to be reproducible.

The issue of reliability in performance analysis has recently been highlighted with respect to the appropriate statistical method for establishing agreement between observations (O'Donoghue, 2007). As the current recommendations are disparate and cast doubt on which existing performance analysis models are indeed reliable, and which technique should be employed with new performance analysis templates, there is a need for consensus and standardisation. In this vein, the statistical approach described by Cooper, Hughes, O'Donoghue and Nevill (2007) has virtue in that it is relatively simple to comprehend and suitable for much of the data recorded in performance analysis, which typically do not lend themselves to parametric statistical techniques (Hughes, Cooper & Nevill, 2002; Nevill, Atkinson, Hughes & Cooper, 2002; Choi et al., 2007; James, Taylor & Stanley, 2007). Cooper et al. (2007) advocate a method which incorporates the non-parametric treatment of test-retest data (Bland & Altman, 1999) and the recommendations of Nevill, Lane, Kilgour, Bowes and Whyte (2001) that 95% of the observed differences should be recorded within a reference value thought to be of 'no practical importance'. This latter point is imperative as it

requires analysts to be knowledgeable about the sport under examination and be able to come to a decision beforehand about how large the test-retest differences in the observations of their performance indicators need to be before they are considered 'important'. In effect, the decision on whether the analysis of the performance indicators is reliable is not dependent upon a statistic that is above or below an arbitrary value, rather it is based on how many events are observed repeatedly within pre-defined limits of acceptability (given that perfect agreement between test and retest is the analyst's goal).

Cooper et al.'s tutorial focused on numerous performance indicators of a particular sport (rugby union) and demonstrated that their technique was sensitive to the level of expertise of the analyst. That is, a less experienced analyst was shown to be less reliable than someone with more experience. Whilst it was argued that the Cooper et al. method was applicable to the field of performance analysis generally, it appears it has yet to be applied to a scenario other than that original one described. Therefore, the aims of the study were to (i) present a notational analysis system for the assessment of the movement characteristics associated with an amateur boxing contest, and (ii) assess, in the manner of Cooper et al. (2007), the reliability of a performance analyst (operator) employing the system.

2. Methods

2.1 Classification of performance indicators

A boxer's performance in a given contest can be appraised simplistically in terms of whether it yielded a victory or a defeat, or more quantitatively in terms of the number of points accrued over the duration of a contest. Notwithstanding the significance of winning by stopping or knocking out an opponent, the events (actions) that lead to the awarding of points by the judges provide the justification and material for performance analysis. Such actions can be described in a typology that defines nine offensive and 12 defensive movements, and four feinting actions (see Tables 1-3,

below), some of which can be identified as occurring in isolation and others in combinations. The quality of such actions can be noted with reference to their intended targets (on the opponent's body) and their outcomes (successful, partially successful or unsuccessful). The lead author, an experienced amateur boxer (25 previous contests) and coach within the sport (> 3 years) initially identified the performance indicators that influence a successful or unsuccessful performance and provided operational definitions for each. The validity of this process was strengthened via consultation with a senior level ABAE coach and another experienced amateur boxer (25 previous contests).

In the context of this study, such performance outcomes were determined visually during post-fight video analysis. Offensive actions (Table 1) were deemed as 'successful' when the attack/punch made visible contact with the opponent's target area, 'partially successful' when the attack/punch was partially blocked or deflected, and 'unsuccessful' when the attack/punch failed to make contact with the opponent's target area. Although it was not possible to corroborate these outcomes with the judges' points allocations, it was reasonable to presume that both 'unsuccessful' and 'partially successful' attacks/punches would not have yielded points, whereas 'successful' attacks/punches, in meeting the ABAE's (2009) criteria for the awarding of points, would. Therefore, actions deemed to be 'successful' may differentiate a victory from a loss. Defensive variables (Table 2) were identified in the same manner. A 'successful' defence resulted in the attack/punch failing to land upon the target area, 'partially successful' defences led to the attack/punch landing upon the defendant after being initially blocked/avoided, and an 'unsuccessful' defence resulted in the attack/punch making visible contact with the defendant's target area despite his attempts to avoid it. Moreover, 'successful' defences and 'partially successful' defences are unlikely to alter the contest score, whereas 'unsuccessful' defences might facilitate the awarding of a point to the aggressor. Feinting motions (potentially less crucial to the outcome of a contest) are described in Table 3 The lead and rear hands were contingent upon the stance adopted by the boxer. That is, boxers adopting an

'orthodox' stance have the left hand as the lead and the right as the rear hand; 'southpaw' stances are the opposite.

2.2 Design

A 3 x 2 minute contest involving two male senior competitors (Light Middleweight, 67 – 71 kg) was chosen at random from a sample of contests (n = 42) recorded as part of on-going research. The contest was recorded with two digital cameras (Canon MV700, Japan) from two adjacent sides of a square ring (4.88 m^2). . The boxers were a 23 year old (24 previous contests, classed as a 'novice'; boxer 'A') and a 21 year old (45 previous contests, classed as an 'open' boxer; boxer 'B'). Performance analysis was conducted post-contest and generally viewed at one quarter of normal playback speed (12.5 frames per second). If necessary, the analyst was permitted to rewind the contest and watch events frame-by-frame. This was justified given the number of actions to notate (25), the speed and complexity of certain movement patterns, particularly those involving combinations, and the desire to capture accurately their outcomes. The two camera angles were used interchangeably, depending upon the location and positioning of the boxers and the referee. The captured data were transferred to a personal computer and subsequently analysed using the Dartfish TeamPro software (version 4.0, Switzerland).

2.3 Contest analysis

For each boxer separately, the events were 'tagged' via the bespoke template (Figure 1) in a sequential manner (Figure 2), commencing with the offensive actions (Table 1) and feints (Table 3), followed by the defensive actions (Table 2). For each strategic offence observed, the overall target and outcome was identified, along with the total number of punches thrown. Thereafter, each individual punch within the attack was coded separately and similarly labelled with its target and outcome. This process was repeated for each strategic defence observed, a difference being the total number of punches *defended*. Where necessary, the analyst was permitted to code multiple,

individual defences simultaneously, regardless of the number of oncoming punches. Additional actions or events occurring in the contest were also notated; the round and its duration, the round and time at which the referee stopped the contest (to issue a warning for example), warnings issued by the referee (for ducking below waist line, excessive holding of an opponent, hitting while holding, dissent and miscellany), eight-second counts issued and the manner in which the contest was won (points verdict, referee stopped contest and knockout).

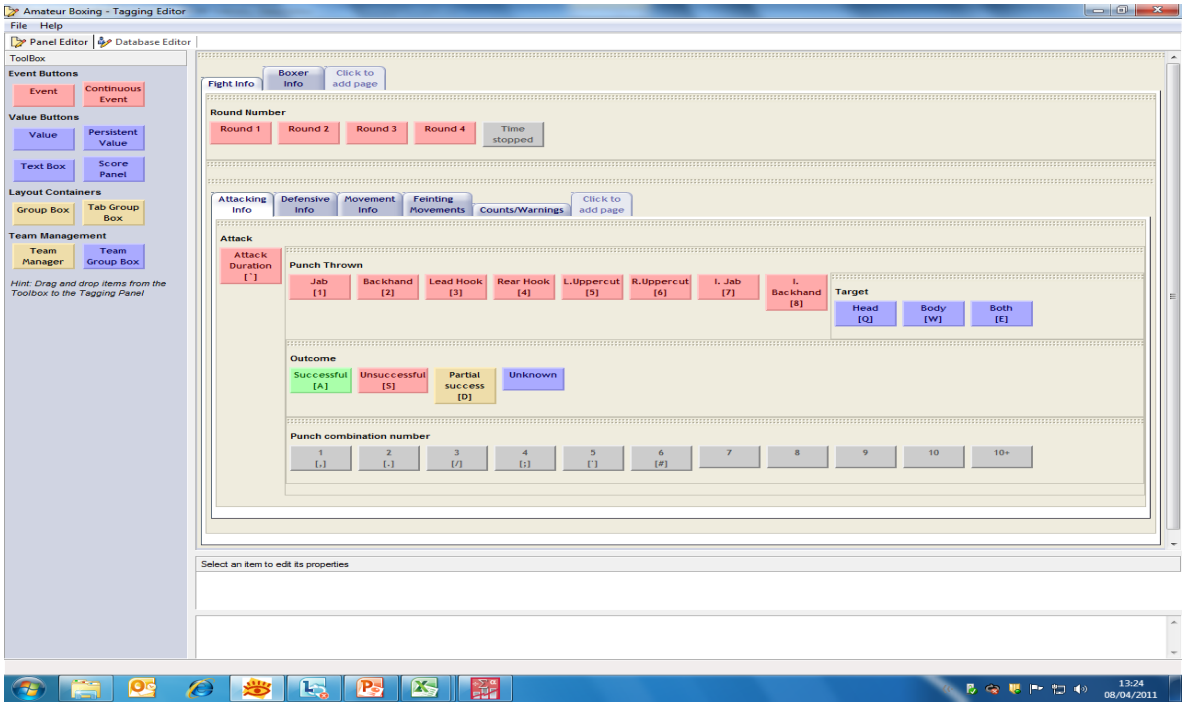

Figure 1. Dartfish analysis template for the coding of offensive behaviours.

Figure 2. A schematic representation of how offensive actions were recorded.

Table 1: Offensive actions.*

Types of attack	Definition
Attack	Any punch or combination of punches performed by a boxer. This indicator is a continuous event in that the duration of the attack is recorded. A^1,A^2,A^3
Jab	A straight punch from the lead hand that moves along the sagittal plane (the central visual line) from anterior to posterior. A^1,A^2

Backhand cross	A straight punch from the rear hand that moves along the sagittal plane (the central visual line) from anterior to posterior. A^1,A^2
Lead hook	A punch from the lead hand that moves along the transverse axis in a sideward 'sweeping' motion. A^1,A^2
Rear hook	A punch from the rear hand that moves along the transverse axis in a sideward 'sweeping' motion. A^1,A^2
Lead uppercut	A punch from the lead hand that moves along the sagittal plane and the longitudinal axis beginning with a downward projection and ending with an upward projection. A^1,A^2
Rear uppercut	A punch from the rear hand that moves along the sagittal plane and the longitudinal axis beginning with a downward projection and ending with an upward projection. A^1,A^2
Inverted jab	A straight punch from the lead hand that moves along the sagittal plane (the central visual line) from anterior to posterior with the arm in a supinated position when extended (palm facing upwards when arm is extended). A^1,A^2
Inverted backhand cross	A straight punch from the rear hand that moves along the sagittal plane (the central visual line) from anterior to posterior with the arm in a supinated position when extended (palm facing upwards when arm is extended). A^1,A^2
Successful attack/ punch (A^1)	A punch is labelled successful when it visibly lands on the opponent's target area. The punch must land directly with the knuckle part of a closed glove on any part of the front or sides of the head or body above the belt line of the opponent. For an attack to be labelled as such, at least one punch must be deemed successful.
Unsuccessful attack/ punch (A^1)	A punch is labelled unsuccessful when it visibly fails to land on the opponent's target area. For example, the punch may land clearly on the arms of the opponent or completely miss the opponent. For an attack to be labelled as such no punches must be labelled as successful or partially successful.
Partially successful attack/ punch (A^1)	A punch is labelled to have partial success when it is partially blocked or deflected yet still lands on the opponent's target area making a visible impact. That is, the punch landed is not a clean punch. For example, a punch may partially land on the arm of an opponent yet still make some form of contact with the opponent's target area. For an attack to be labelled as such no punch should be deemed successful yet at least one punch should be deemed as partially successful.
Head (A^2)	A punch is labelled as being aimed towards the head if it visibly lands on the opponent's head or misses the head of the opponent.
Body (A^2)	A punch is labelled as being aimed towards the body if it visibly lands on the opponent's body or misses the body of the opponent.
Both (A^2)	Only attacks can be labelled as such. An attack is labelled as being aimed towards 'both' when the combination of punches involves at least one punch aimed towards the head and one punch towards the body.
Attack combination number (A^3)	Only attacks are labelled with this. Labelled as the number of punches involved in that attack.

* Each action was labelled with respect to its outcome (A^1), target (A^2) and combination (A^3) (as described in the shaded areas)

Table 2: Defensive actions*.

Types of Defence	**Definition**
Defence	Any defence/ combination of defences performed by a boxer. This indicator is a continuous event in that the duration of the attack is recorded. D^1, D^2, D^3

Slip left	Movement of the head and/or trunk to the left in order to avoid a punch. D^1, D^2
Slip right	Movement of the head and/or trunk to the right in order to avoid a punch. D^1, D^2
Lean backwards	Movement of the head and/or trunk and/or flexion of the rear leg leaning the boxer's target area (predominantly the head) away from the attacker in order to avoid a punch. D^1, D^2
Duck	Movement achieved by flexion of the knee joints and/or trunk in order to lower the boxer's target area (predominantly the head) in order to avoid a punch. D^1, D^2
Role clockwise	Movement of the head and trunk whereby the boxer's target area (predominantly the head) is moved in a circular motion beginning with movement to the left. D^1, D^2
Role anti-clockwise	Movement of the head and trunk whereby the boxer's target area (predominantly the head) is moved in a circular motion beginning with movement to the right. D^1, D^2
Block/parry with lead arm	Movement of the lead arm whereby it deflects an oncoming punch away from the target area or placement of the arm over the target area so the punch lands on the arm instead of the target area. D^1, D^2
Block/parry with rear arm	Movement of the rear arm whereby it deflects an oncoming punch away from the target area or placement of the arm over the target area so the punch lands on the arm instead of the target area. D^1, D^2
Block both arms	Movement of both arms whereby the arms are positioned in a manner that attempts to cover the boxer's own target area so that the punch lands on the arm instead of the target area. D^1, D^2
Foot defence	Movement whereby the boxer transports his centre of mass away from the attacker to avoid punches directed towards them. D^1, D^2
Clinch	Movement whereby a boxer holds an opponent's body and/or arms with one or both of his arms to prevent or hinder the opponent's punches or movements. D^1, D^2
Successful defence (D^1)	A defence is deemed successful if it led to the punch missing the target area or failing to visibly land on the target area.
Partially successful defence (D^1)	A defence is deemed partially successful if it the oncoming punch or punches initially blocked or avoided yet still made some form of contact with the defendants target area.
Unsuccessful defence (D^1)	A defence is deemed unsuccessful if it failed to prevent the punch landing on the target area.
Head (D^2)	A defence is labelled as such if it was performed in order to protect the individual's head.
Body (D^2)	A defence is labelled as such if it was performed in order to protect the individual's head.
Both (D^2)	A defence is labelled as such if it was performed in order to protect the individual's body.
Defence combination number (D^3)	Only defences are labelled as such. Labelled as the number of punches defended against.

* Each action was labelled with respect to its outcome, target, and combination (as described in the shaded areas)

Table 3: Performance actions relating to feinting movements. .

Types of feint	**Definition**

Foot feint	A quick movement of the feet, usually a short shuffle forwards, that is performed with the intention of misleading the opponent into believing he is going to move towards the opponent.
Head/Body feint	A quick movement of the head or trunk that is performed with the intention of misleading the opponent into believing he is possibly going to launch an attack on the opponent.
Lead hand feint	A quick movement of the lead hand, replicating the initiation of a lead hand punch, which is performed with the intention of misleading the opponent into believing he is e going to throw a lead hand punch.
Rear hand feint	A quick movement of the rear hand, replicating the initiation of a rear hand punch, which is performed with the intention of misleading the opponent into believing he is going to throw a rear hand punch.

2.4 Intra- and inter-observer reliability analysis

The full contest (three rounds) was analysed on two occasions four weeks apart by the lead author and subjected to intra-observer reliability analysis. Subsequently, his first round data (initial analysis) was used as a reference against which the performances of two other observers were compared, thereby enabling an assessment of the inter-observer reliability (agreement). The two observers were an amateur boxer (AB; 25 previous contests) who had no prior experience of performance analysis but was also an experienced boxing coach, and a knowledgeable performance analyst, though not previously of boxing. On different occasions, each individual was given the operational definitions of the performance indicators to read before being exposed to the test data in the Dartfish programme. Where necessary, clips of example boxing footage were shown to aid their understanding of the performance indicators.

2.5 Statistical analyses

The method proposed by Cooper et al. (2007) was used to quantify the intra- and inter-operator reliability of the performance analysis model described above. Whilst the reader is referred to their article for an in-depth explanation of this methodology, it originates from Bland and Altman's

(1999) paper on assessing agreement when the distributions of the data do not satisfy the assumption of normality. The reliability statistics generated were for each boxer individually and likewise for each performance indicator.

A feature of the methodology proposed by Cooper et al. (2007) was their division of a selected sport performance (an 80-minute rugby union match) into discrete two-minute time cells, yielding approximately 40 cells (depending on the amount of over-time played) of data. This 'sample' of data was deemed sufficient to enable a worthwhile test-retest analysis in the absence of access to a large number of separate matches and the greater amount of time needed to analyse them. It was posited that for the performance indicators chosen (e.g. numbers of passes and tackles), such a time period was appropriate due to their relatively frequent occurrences and, implicitly, that there would be few, if any, 'empty' cells. Arguably, therefore, longer time cells would suit the analysis of infrequent events and/or longer sports performances (e.g. a three-day cricket match), and shorter ones for the analysis of rapidly occurring events and/or shorter performances (e.g. a boxing contest). Accordingly, a 10 s time cell (12 per round, up to 36 per bout) was chosen for the current study.

A median sign test was computed to assess the null hypothesis of no significant systematic bias between the test and retest scores (frequency counts) of each action. Subsequently, the observed proportion of agreement was calculated. This involved the *a priori* determination of the proportion of differences that was greater than some reference value deemed to be of no 'practical importance' (Nevill et al., 2001). Somewhat arbitrarily, Cooper et al (2007) selected a reference value of ±1 (actions) for their rugby data, but they acknowledged that the type and frequency of the performance data would have a bearing on the choice of this value. In the case of an amateur boxing contest, many offensive actions (punches) and defences are performed during a bout (e.g. >112 punches per round during a 3 x 3 minute contest; Smith et al., 2001) with the chances of a knockout

blow resulting from a single successful attack/punch or unsuccessful defence being relatively small. Furthermore, the final number of points awarded to competitors is often less than 10 (European Boxing Confederation, 2011), implying that the frequency of specific point-yielding actions is low. On this basis, a judgement was made that the boxing analyst should strive for a narrow reference range (margin of error), in order to minimise the likelihood of missing one of the few, pivotal actions in a round/bout. Accordingly, Cooper at al.'s reference value of ±1 seemed appropriate in this context, along with a target of proportion of total agreement of ≥ 95%.

The degree of *perfect* agreement, *p*o, was calculated for each indicator as the correctly observed proportion (*r*) out of the total observed number (*n*) of the test-retest scores entered ($po = r/n$), along with the degree of agreement within the reference value of ±1. Approximate confidence intervals were then calculated for these proportions of agreement (upper 95% CI = po + (1.96 x $SE(po)$); lower 95% CI = po - (1.96 x $SE(po)$). The results described below pertain to boxer A, and unless indicated otherwise, can be assumed to be very similar to those for boxer B.

3. Results

3.1 Intra-observer agreement

There were no significant differences ($P > 0.05$) between the analyst's test and retest observations for all the performance indicators (Table 4). The proportion of total agreement (PA) ranged between 92 – 100% and when the reference value of ± 1 was considered (PA ± 1), agreement was 100% for all indicators. When the outcome of each particular action (its 'success') of boxer A was considered separately to its mere occurrence, the proportion of total agreement was often 100%, and no less than 92%. For PA ± 1, the agreement was 100% for all actions.

Table 4: Summarised intra-observer test-retest values – boxer A.

Performance indicator	Median (sign test)	PA = 0 (%)	95% Confidence Interval (%)	PA ± 1 (%)	95% Confidence Interval (%)
Attack	*P* = 1.00	100	100 to 100	100	100 to 100
Jab	*P* = 1.00	100	100 to 100	100	100 to 100
Backhand	*P* = 1.00	95	87 to 100	100	100 to 100
Lead hook	*P* = 1.00	97	92 to 100	100	100 to 100
Rear hook	*P* = 1.00	95	87 to 100	100	100 to 100
Lead uppercut	*P* = 1.00	97	92 to 100	100	100 to 100
Rear uppercut	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted jab	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted backhand	*P* = 1.00	100	100 to 100	100	100 to 100
Lead hand feint	*P* = 1.00	100	100 to 100	100	100 to 100
Rear hand feint	*P* = 1.00	100	100 to 100	100	100 to 100
Head/Body Feint	*P* = 0.25	95	87 to 100	100	100 to 100
Foot feint	*P* = 1.00	92	83 to 100	100	100 to 100
Defence	*P* = 1.00	97	92 to 100	100	100 to 100
Block both arms	*P* = 1.00	100	100 to 100	100	100 to 100
Block right arm	*P* = 1.00	97	92 to 100	100	100 to 100
Block left arm	*P* = 1.00	100	100 to 100	100	100 to 100
Clinch	*P* = 1.00	100	100 to 100	100	100 to 100
Duck	*P* = 1.00	100	100 to 100	100	100 to 100
Foot defence	*P* = 1.00	97	92 to 100	100	100 to 100
Lean back	*P* = 1.00	97	92 to 100	100	100 to 100
Push	*P* = 1.00	100	100 to 100	100	100 to 100
Slip left	*P* = 1.00	100	100 to 100	100	100 to 100
Slip right	*P* = 1.00	100	100 to 100	100	100 to 100
Roll clock	*P* = 1.00	100	100 to 100	100	100 to 100
Roll anti-clockwise	*P* = 1.00	100	100 to 100	100	100 to 100

Key: PA = proportion of total agreement; PA ± 1 = proportion of agreement within the reference value of ± 1; N/A = not applicable.

An example of the agreement for an offensive and defensive indicator (the backhand and block with the right arm) across the 37 10 s time cells of the bout is presented in Table 5.

Tables 5 and 6 serve to illustrate the non-parametric method for determining the reliability of test-retest data for four performance indicators. The PA for the backhand (Table 5) was = 0.95, or 95% (35/37) since though a total of 27 backhands were recorded in both observations, indicating a reliable analysis, perfect agreement was not established as the analyst failed to record the same number of backhands during time cells 30 and 37. For the block with right arm, 36 time cells agreed, with only a single error occurring in time cell 34, yielding a PA of 97%. For the example of a frequently occurring action (attack), perfect reliability (100%) was reflected by the total (71 instances) and consistent recordings across all time cells (Table 6). For the infrequent action (lead uppercuts), four actions were recorded during the initial analysis and five during the retest, but agreement occurred for 36/37 time cells, yielding a PA of 97%.

Table 5: Intra-observer reliability data for an offensive (backhands) and defensive action (block with the right arm) recorded by the expert analyst within the 10 s time cells for boxer A.

Cell number*	Backhand	Backhand retest	Backhand: same data in test retest	Block right arm	Block right arm retest	Block right arm: same data in test retest
1	0	0	Yes	1	1	Yes
2	3	3	Yes	1	1	Yes
3	0	0	Yes	0	0	Yes
4	0	0	Yes	1	1	Yes
5	1	1	Yes	1	1	Yes
6	1	1	Yes	0	0	Yes
7	0	0	Yes	1	1	Yes
8	0	0	Yes	0	0	Yes
9	0	0	Yes	0	0	Yes
10	0	0	Yes	0	0	Yes
11	1	1	Yes	0	0	Yes
12	0	0	Yes	2	2	Yes
13	1	1	Yes	4	4	Yes
14	0	0	Yes	1	1	Yes
15	1	1	Yes	2	2	Yes
16	1	1	Yes	0	0	Yes
17	0	0	Yes	0	0	Yes
18	0	0	Yes	1	1	Yes
19	0	0	Yes	0	0	Yes
20	1	1	Yes	0	0	Yes
21	3	3	Yes	0	0	Yes
22	0	0	Yes	1	1	Yes
23	1	1	Yes	1	1	Yes
24	1	1	Yes	2	2	Yes
25	1	1	Yes	1	1	Yes
26	1	1	Yes	2	2	Yes
27	0	0	Yes	1	1	Yes
28	2	2	Yes	2	2	Yes
29	0	0	Yes	1	1	Yes
30	0	1	No	0	0	Yes
31	0	0	Yes	1	1	Yes
32	0	0	Yes	1	1	Yes
33	1	1	Yes	0	0	Yes
34	1	1	Yes	0	1	No
35	3	3	Yes	0	0	Yes

36	1	1	Yes	1	1	Yes
37	2	1	No	0	0	Yes
Total	**27**	**27**	**Yes = 35** **No = 2**	**29**	**30**	**Yes = 36** **No = 1**

***37 times cells since round 1 exceeded 2 minutes**

Table 6: Intra-observer reliability data for a frequent (attack) and infrequent action (lead uppercut) recorded by the expert analyst within the 10 s time cells for boxer A.

Cell number	Attack	Attack retest	Attack: same data in test retest	Lead uppercut	Lead uppercut retest	Lead uppercut: same data in test retest
1	1	1	Yes	0	0	Yes
2	2	2	Yes	1	1	Yes
3	2	2	Yes	0	0	Yes
4	1	1	Yes	0	0	Yes
5	3	3	Yes	1	1	Yes
6	3	3	Yes	0	0	Yes
7	1	1	Yes	0	0	Yes
8	1	1	Yes	0	0	Yes
9	2	2	Yes	0	0	Yes
10	2	2	Yes	0	0	Yes
11	3	3	Yes	0	0	Yes
12	1	1	Yes	0	0	Yes
13	3	3	Yes	0	0	Yes
14	2	2	Yes	1	1	Yes
15	1	1	Yes	0	0	Yes
16	4	4	Yes	0	0	Yes
17	2	2	Yes	0	0	Yes
18	2	2	Yes	0	0	Yes
19	1	1	Yes	0	0	Yes
20	2	2	Yes	0	0	Yes
21	2	2	Yes	0	1	No
22	1	1	Yes	0	0	Yes
23	2	2	Yes	0	0	Yes
24	3	3	Yes	0	0	Yes
25	1	1	Yes	0	0	Yes
26	2	2	Yes	0	0	Yes
27	2	2	Yes	0	0	Yes
28	3	3	Yes	0	0	Yes
29	2	2	Yes	0	0	Yes
30	2	2	Yes	0	0	Yes
31	2	2	Yes	0	0	Yes
32	0	0	Yes	0	0	Yes
33	2	2	Yes	1	1	Yes
34	2	2	Yes	0	0	Yes
35	3	3	Yes	0	0	Yes
36	2	2	Yes	0	0	Yes

37	1	1	Yes	0	0	Yes
Total	**71**	**71**	**Yes = 71** **No = 0**	**4**	**5**	**Yes = 36** **No = 1**

***37 times cells since round 1 exceeded 2 minutes**

3.2 Inter-observer agreement

3.2.1 Reference analyst versus AB analyst

The agreement between the analyses of the reference (lead author) and the AB was less impressive than that for the intra-observer reliability analysis, though it is noteworthy that there was no systematic bias between the observers for any performance indicator (Table 7). Moreover, total agreement occurred for the majority of indicators and for all indicators when the ±1 range was considered.

Table 7: Summarised inter-observer test-retest values (reference versus AB analyst) – boxer A

Performance indicator	Median (sign test)	PA = 0 (%)	95% Confidence Interval (%)	PA ± 1 (%)	95% Confidence Interval (%)
Attack	*P* = 1.00	100	100 to 100	100	100 to 100
Jab	*P* = 1.00	100	100 to 100	100	100 to 100
Backhand	*P* = 0.50	83	62 to 104	100	100 to 100
Lead hook	*P* = 1.00	92	76 to 100	100	100 to 100
Rear hook	*P* = 0.50	83	62 to 100	100	100 to 100
Lead uppercut	*P* = 1.00	92	76 to 100	100	100 to 100
Rear uppercut	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted jab	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted backhand	*P* = 1.00	100	100 to 100	100	100 to 100
Lead hand feint	*P* = 1.00	92	76 to 100	100	100 to 100
Rear hand feint	*P* = 1.00	100	100 to 100	100	100 to 100
Head/Body Feint	*P* = 0.50	83	62 to 100	100	100 to 100
Foot feint	*P* = 1.00	100	100 to 100	100	100 to 100
Defence	*P* = 1.00	100	100 to 100	100	100 to 100
Block both arms	*P* = 1.00	92	76 to 100	100	100 to 100
Block right arm	*P* = 1.00	83	62 to 100	100	100 to 100
Block left arm	*P* = 1.00	100	100 to 100	100	100 to 100
Clinch	*P* = 1.00	92	76 to 100	100	100 to 100
Duck	*P* = 1.00	100	100 to 100	100	100 to 100

Foot defence	*P* = 1.00	75	51 to 100	100	100 to 100
Lean back	*P* = 1.00	83	62 to 100	100	100 to 100
Push	*P* = 1.00	100	100 to 100	100	100 to 100
Slip left	*P* = 1.00	100	100 to 100	100	100 to 100
Slip right	*P* = 1.00	100	100 to 100	100	100 to 100
Role clock	*P* = 1.00	100	100 to 100	100	100 to 100
Role anti-clockwise	*P* = 1.00	100	100 to 100	100	100 to 100

Key: PA = proportion of total agreement; PA ± 1 = proportion of agreement within the reference value of ± 1.

3.2.2 Reference analyst versus experienced performance analyst

For all performance indicators there was no systematic bias between analysts (Table 8) and the degree of total agreement was 100% in most cases. Outstanding though was the relatively poor agreement for three feinting actions (50 – 83%). These inconsistencies were not present when the ± 1 value was calculated, except for one (head/body feint) action.

Table 8: Summarised inter-observer test-retest values (reference versus expert performance analyst) – boxer A.

Performance indicator	Median (sign test)	PA = 0 (%)	95% Confidence Interval (%)	PA ± 1 (%)	95% Confidence Interval (%)
Attack	*P* = 1.00	100	100 to 100	100	100 to 100
Jab	*P* = 1.00	92	76 to 100	100	100 to 100
Backhand	*P* = 1.00	92	76 to 100	100	100 to 100
Lead hook	*P* = 1.00	100	100 to 100	100	100 to 100
Rear hook	*P* = 1.00	92	76 to 100	100	100 to 100
Lead uppercut	*P* = 1.00	92	76 to 100	100	100 to 100
Rear uppercut	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted jab	*P* = 1.00	100	100 to 100	100	100 to 100
Inverted backhand	*P* = 1.00	100	100 to 100	100	100 to 100
Lead hand feint	*P* = 0.50	83	62 to 100	100	100 to 100
Rear hand feint	*P* = 1.00	100	100 to 100	100	100 to 100
Head/Body Feint	*P* = 0.69	50	22 to 78	75	51 to 100
Foot feint	*P* = 0.25	75	51 to 100	100	100 to 100
Defence	*P* = 1.00	92	76 to 100	100	100 to 100
Block both arms	*P* = 0.50	83	62 to 100	100	100 to 100
Block right arm	*P* = 1.00	100	100 to 100	100	100 to 100
Block left arm	*P* = 1.00	100	100 to 100	100	100 to 100
Clinch	*P* = 1.00	100	100 to 100	100	100 to 100
Duck	*P* = 1.00	92	76 to 100	100	100 to 100

Foot defence	$P = 1.00$	100	100 to 100	100	100 to 100
Lean back	$P = 1.00$	92	76 to 100	100	100 to 100
Push	$P = 1.00$	100	100 to 100	100	100 to 100
Slip left	$P = 1.00$	100	100 to 100	100	100 to 100
Slip right	$P = 1.00$	92	76 to 100	100	100 to 100
Roll clockwise	$P = 1.00$	100	100 to 100	100	100 to 100
Roll anti-clockwise	$P = 1.00$	100	100 to 100	100	100 to 100

Key: PA = proportion of total agreement; PA ± 1 = proportion of agreement within the reference value of ± 1.

4. Discussion

This paper has presented a unique performance analysis model (template) for amateur boxing and reported on its reliability through intra- and inter-observer comparisons. The template was established through content validity procedures by two experienced amateur boxers with coaching experience and an advanced level amateur boxing coach. This yielded the identification of 25 performance indicators (actions), with assignable values reflecting the intended target and outcome. In its current form the template is designed to be used via video replay post-contest of successive, discrete 10 s cells, and *not* specifically by a highly trained performance analyst.

In adopting an appropriate statistical approach for data of this kind, it emerged that the level of *intra*-observer reliability was excellent, with the test-retest frequency scores (of each time cell) for most indicators demonstrating 100% agreement, and better than 91% agreement across all indicators. When the pre-specified tolerance zone (reference value) of ± 1 counts was considered, all the performance indicators were notated accurately over the repeat trials. For the *inter*-observer analysis, the degree of perfect agreement was lower than intra-observer, but was nevertheless excellent for both the AB and the expert analyst, with all but one indicator showing 100% agreement within the reference value of ± 1. Indeed, the level of *perfect* agreement was often 100%, though it was only 50% for a particular feinting action. It is clear that given adequate familiarisation with the performance template, an amateur boxing contest (filmed from at least two camera angles)

can be reliably notated by individuals neither particularly experienced in boxing nor in performance analysis.

That the level of inter-observer reliability was somewhat inferior to the intra-observer reliability was not unexpected and has been observed previously during team game analysis (James, Mellalieu & Hollely, 2002; James, Mellalieu & Jones, 2005; Tenga et al., 2009; Worsfold & Macbeth, 2009). It is plausible that this could be due simply to the observer's lack of familiarity with the analysis template and/or the sport of boxing, or a degree of imprecision in the operational definitions of the performance indicators. In the case of the latter, as the actions are performed in a very dynamic environment, any disparity between the numbers of observations was likely due to events being misclassified, rather than not being coded at all. An example of this occurred when the amateur boxer coded two events as rear hooks whereas the reference analyst coded them as backhands, producing four errors. Now whilst the operational definitions should be clear enough to distinguish between these two different punches, in certain situations they share many characteristics, making it very difficult to distinguish between them. Such an incidence is recognised as a recurrent problem in performance analysis (Hughes, Cooper, Nevill & Brown, 2003). Moreover, for certain indicators, the dynamic nature of the contest alone will inevitably lead to some errors both between, and within observers (Hughes et al., 2002; James et al., 2007).

Similar levels of reliability were seen in the two inter-observer conditions across most performance indicators, and this demonstrates that the use of the template does not require expert knowledge of the sport's actions or expertise in performance analysis. This is probably because most of the actions identified are fundamental, gross movements that are easy to discern and notate. However, in the case of the feinting actions, the expert performance analyst was less adept at identifying their occurrence than the lead author and AB, suggesting an increased knowledge of the sport, with regard to the feinting actions, may be advantageous. For such actions, with their subtle

characteristics, the operational definitions might need refining or the analyst needs more exposure to them. Additional modifications to the system described in this paper would be necessary, given its time-consuming post-event procedure, if the goal was to analyse boxing contests in real-time (O'Donoghue, 2008). Conversely, the analysis template could be expanded to incorporate the foot movement/orientation of the boxers around the ring and provide a more comprehensive profile of an individual's performance. Nonetheless, the current system has the potential to elucidate the characteristics and demands of amateur boxing and inform the training and competitive practices of its competitors.

5. Conclusions

This study has demonstrated that a novel performance analysis template can yield consistent (reliable) observations of the key movement characteristics occurring in a pre-recorded amateur boxing bout. Importantly, where a reference or 'error' limit ± 1 is set, the template can be used reliably by different operators, having varying experiences of performance analysis. Whilst the nature of the current template (in terms of the number and type of actions recorded, and their outcomes) has rendered the process a rather lengthy one, the depth of the analysis provides the basis for scrutiny by coaches seeking to identify specific markers of successful performances. Potentially, the template could be streamlined to facilitate a more rapid performance analysis, and indeed be readily adapted for the professional version of the sport. Moreover, the template has enabled the identification of the movement characteristics of typical boxing bouts that, in the contemporary manner of other sports, could be transposed into a simulation protocol for the purpose of administering boxing-specific conditioning and monitoring the effects of performance-enhancing interventions.