European Journal of Human Movement, 2017: **39**, 65-81

GENERALIZABILITY THEORY APPLIED TO OLYMPIC MALE TAEKWONDO COMBATS

Cristina Menescardi [1]; Isaac Estevan [2]; Coral Falco [3]; Antonio Hernández-Mendo [4]

1. Department of Motricity and Teaching of Physical Activity, Catholic University of Valencia "San Vicente Mártir", Valencia, Spain.
2. Department of Teaching of Music, Visual and Corporal Expression, University of Valencia, Valencia, Spain.
3. Department of Health Promotion and Development, University of Bergen, Bergen, Norway.
4. Department of Social Psychology, Social Work, Anthropology and East Asian Studies, University of Málaga, Spain.

ABSTRACT

Introduction: Generalizability theory (GT) postulates that there are infinite sources of error variation in any measurement. Generalizability analysis is used to reduce and control this measurement error. To date, there has been no study employing GT in analysing taekwondo, despite the increase of observational studies in combat sports. Through applying GT to six random combat bouts within the Olympic Games in London 2012, the aims of this study are: (a) to determine whether the bouts selected are representative enough for their behaviour to be generalizable; (b) to determine the intra and interobserver agreement and reliability; and (c) to assess the accuracy of the exhaustive and mutual exclusion of each category included in each criterion of the categorical system. Method: Six taekwondo bouts were randomly selected in order to analyse the athletes' technical/tactical behaviour. Three observers were responsible for viewing and analysing the videos using a consensual concordance strategy. Results and discussion: High values for intra and interobserver agreement were found in Kendall's tau and Cohen's kappa indexes, in addition to the GT. The results showed that the categorical system was composed of exhaustive and mutually exclusive (E/ME) categories and that the sample was representative enough to generalize the behaviour shown.

Key words: taekwondo athletes, validation, observational tool, observational methodology

TEORÍA DE LA GENERALIZABILIDAD APLICADA A LOS COMBATES MASCULINOS DE TAEKWONDO OLÍMPICO

RESUMEN

Introducción: La Teoría de la Generalizabilidad (TG) postula que hay infinitas fuentes de variación en cada medición. El análisis de generalizabilidad se utiliza para reducir y controlar este error de medición. Hasta la fecha, no se han encontrado estudios que hayan empleado la TG en taekwondo, a pesar de que haya incrementado el número de estudios observacionales en los deportes de combate. Mediante el empleo de la TG, a seis combates seleccionados al azar de los Juegos Olímpicos de Londres 2012, el objetivo del estudio es: (a) determinar si los combates seleccionados son suficientemente representativos para generalizar el comportamiento; (b) determinar el acuerdo y fiabilidad intra e interobservador; y (c) evaluar la precisión de la exhaustividad y mutua exclusividad de las categorías incluidas en cada criterio del sistema de categorías. Método: Se seleccionaron seis combates de taekwondo para analizar el comportamiento técnico-táctico de los deportistas. Tres observadores fueron responsables del visionado y de codificar los vídeos utilizando la estrategia de concordancia consensuada. Resultados y discusión: Se encontraron altos valores de acuerdo intra e interobservador en los índices tau de Kendall y kappa de Cohen, además de en la TG. Los resultados mostraron que el sistema de categorías estaba compuesto por categorías exhaustivas y mutuamente excluyentes (E/ME) y que la muestra es suficientemente representativa para generalizar el comportamiento mostrado por los deportistas.

Palabras clave: taekwondistas, validación, herramienta observacional, metodología observacional

Correspondence:
Cristina Menescardi
Department of Motricity and Teaching of Physical Activity, Catholic University of Valencia "San Vicente Mártir", Valencia, Spain.
ca.menescardi@ucv.es
Submitted: 27/02/2017 *Accepted:* 12/12/2017

INTRODUCTION

Studies which analyse sporting behaviour are dynamic and must therefore reflect the reality of the sport and its competition regulations. Taekwondo is an Olympic combat sport whose rules have changed much over the years. In this sport, athletes mainly perform leg techniques (Kazemi, Casella, & Perri, 2009), seeking to achieve a higher score than the opponent and thus win the bout (WTF, 2012). During the Olympic Games in London in 2012, points were achieved by: (a) kicks to the trunk of the body, scoring one point; or (b) kicks to the head, scoring three points. Moreover, if the technique was performed using a spin beforehand, an additional point was added (WTF, 2012). With regard to the nature of the bout, the scientific literature indicates that depending on how the authors perceived performance in this sport there are various relevant criteria which can affect the athlete's behaviour in terms of the tactical action taken, such as the type of technique performed, the height of the target, the laterality, or the effectiveness of technical-tactical actions (Kazemi et al., 2009; Falco, Landeo, Menescardi, Bermejo, & Estevan, 2012; Falco, Estevan, Alvarez Morales-Sánchez, & Hernández-Mendo, 2014). One valid way to measure the performance criteria is through the use of observational methodology (Yiannakos & Armatas, 2006). In view of the above performance criteria, several observational studies have been conducted using various observational tools according to the purposes of the researchers, such as technical analysis (Tornello et al., 2014), tactical analysis (Kazemi et al., 2009) or both (Falco et al., 2012, 2014).

The observational tools used in these studies were characterized for validity by consultation with experts, and this approach allows an assessment of the content validity from a qualitative perspective (Mérida, Serrano, & Tabernero, 2015). However, a quantitative perspective is also required in order to analyse the reliability of the observers and the data obtained, in addition to the adequacy of the observational tool (construct validity). Since there are numerous factors which can affect an athlete's sporting performance, the development and validation of an observational tool for Olympic-level taekwondo is necessary. This observational tool should be based on previous studies and should involve an improved design by taking into consideration all the salient performance variables, as well as ensuring reliable, valid, and accurate observations; it should also be generalizable to populations with similar characteristics to those in the sample. The validation of such an observation system can be carried out using Generalizability Theory (GT) (Blanco-Villaseñor, Castellano, Hernández-Mendo, Sánchez-López, & Usabiaga, 2014). This theory has been considered from several perspectives, for example as (a) an extension of Classical Test Theory, using an analysis involving variance methods and experimental designs (Martínez, 1996); and (b) as a

multifaceted error theory, which assumes infinite sources of variation (facets) in any measurement situation (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). GT therefore allows for the analysis of numerous facets that may be affecting a measure or a measurement design of observational origin.

In this sense, GT considers two factors of particular relevance (Blanco-Villaseñor, Castellano, & Hernández-Mendo, 2000): (a) the number of observers encoding behaviours and (b) the observational tool used. With regard to the number of observers, the scientific literature reflects the importance of including more than one, so that the consistency of the data is not affected by the subjective judgment of a single observer (Medina & Delgado, 1999). However, including more than one observer can lead to differences in coding between them. To increase the reliability of the data obtained, the use of interobserver agreement (when two observers encode the same observation) and intraobserver agreement (when the observer encodes the same observation twice) is suggested. These values will indicate the extent to which the data has remained stable throughout the process of encoding (Medina & Delgado, 1999; Blanco-Villaseñor et al., 2014). In addition, the observational tool created must be precise in defining the behaviours observed. A behaviour cannot be classified in two categories of the observational tool simultaneously; that is, each category (a concrete action within specific sport) must be exhaustive and mutually exclusive (E/ME), at least for each criterion (the type of behaviour classified according to the categories of the observation tool) (Anguera, Blanco, Hernández-Mendo, & Losada, 2011; Blanco-Villaseñor et al., 2014). In this sense, the term 'exhaustive' refers to the inclusion of all behaviours under study (within certain criterion), which can be assigned to one of the categories; while the term 'mutually exclusive' means that each category of the observational tool represents only a concrete action; for this reason, an action cannot be encoded in more than one category. Failure to fulfil the E/ME condition implies poor definition of the categories and the necessity of reviewing the definitions used in the observational tool, since this would mean that the data were unreliable, and possibly even a distortion of reality (Anguera, Magnusson, & Jonsson, 2007).

After verifying the reliability and accuracy of the measurement design, GT allows optimization in terms of costs and benefits; the whole sample need not be analysed since the results may be generalized from a random sample of the target population (McKenzie, 2010). Previous studies in the field of sports have shown generalizable data from the analysis of between one and 12 matches, depending on the tournament or league and assuming that at least 5%–10% of the population needs to be analysed (Blanco-Villaseñor et al., 2000; Usabiaga, Castellano, Blanco-Villaseñor, & Casamichana, 2013). The lack of studies using GT in taekwondo forms the motivation for the current work. Thus, by applying

GT to six selected randomly bouts of the Olympic Games in London in 2012, the purposes of this study were: (a) to determine whether the number of bouts selected allows the generalization of these athletes' behaviour; (b) to determine the inter and intraobserver agreement and reliability; and (c) to assess the accuracy of the E/ME categories used in each of the criteria of the observational tool, in order to obtain reliable data.

METHOD

Participants

Six male bouts of the Olympic Games in London 2012 were simple randomly selected and analysed (≤58 kg = one final; 58–68 kg = one final and one quarter-final; 68–80 kg = one final and one quarter-final and ≥80 kg = one semi-final). To be considered on the analysis, the duration of the bouts had to be three rounds, according to the observational methodology requirements (intra and intersessional duration). According to *The Belmont Report* (1978), since the analysed videotapes are in the public domain and public behaviour is shown, it is not necessary to acquire informed consent from the athletes observed.

Procedure

To analyse all the bouts, an *ad hoc* observational tool was created using a combination of the categorical system and format fields. The definitions of the 11 criteria (Table 1) forming the observational tool (round, weight category, event, colour protector, tactical action, type of technique, height target, laterality, kicking leg, guard or score), were developed taking into account the background of the martial artists' tactical behaviour as described by previous studies (Kazemi et al., 2009; Casolino et al., 2012; Falco et al., 2012, 2014; Menescardi, López-López, Falco, Hernández-Mendo, & Estevan, 2015). In this tool, the observational unit used was the combat action, in which one or both contestants try to score or avoid being scored against. Only kicking actions were considered for this analysis, since this type is most frequently used in the competition (Kazemi et al., 2009). Of arm actions, only hand-blocks were considered in the analysis for their relevance to the kicking defence. Each action (category) was defined according to a categorical core in terms of its main content. Following the recommendations made by Anguera et al. (2007), a passive observation process followed by several meetings were held separately with two taekwondo experts in order to develop the categories; two experts in observational methodology were also consulted in order to assess the correct adjustment of the observational tool.

TABLE 1
Categorical system of the observational tool.

CRITERION	TYPE	CATEGORY	CATEGORICAL CORE
Round	Mixed	First	First round of the bout
		Second	Second round of the bout
		Third	Third round of the bout
Weight category	Fixed	Fly	Weight category ≤ 58kg
		Feather	Weight category between (58 - 68] kg
		Welter	Weight category between (68 - 80] kg
		Heavy	Weight category ≥ 80 kg
Event	Fixed	Preliminary	First phase of the championship
		Quarter final	Second phase of the championship
		Repechage	Bout for determining bronze medal contestants
		Semi final	Third phase of the championship
		Bronze medal	Consolation final between the two third-placed contestants
		Final	Fourth and final phase of the championship
Colour protector	Fixed	Blue	The colour of the competitor's trunk protector is blue
		Red	The colour of the competitor's trunk protector is red
Tactical action	Changeable	Block	Defensive actions to avoid the impact of a kick by placing one arm or leg between the protector and the leg of the opponent. This does not have a scoring objective
		Dodge	Defensive movement to avoid being kicked by the opponent. This does not have a scoring objective
		Cut	Defensive forward movement to avoid being beaten by a close opponent, and to prevent the attacking action from being completed. This does not have a scoring objective
		Opening	Movement to control the distance with the opponent or bridge the gap between both competitors
		Direct attack	Offensive action with the objective of scoring, ending with an impact on the opponent but without previous movement
		Indirect attack	Offensive action in order to score, ending with an impact on the opponent and with previous movement such as a step, skip, opening, guard change, kicking trajectory modification, etc.
		Anticipated counterattack	Action that starts during the opponent's attack with the purpose of scoring. The athlete kicks the attacker during the preparatory phase (guard) and/or initial phase (when the opponent's knee is being raised)
		Simultaneous counterattack	Action that starts at the same time as the opponent's attack and has a scoring purpose. The athlete kicks at the same time as the opponent. Thus, the counter attacker kicks at the end of the attacker's initial phase (leg raised) or during the impact momentum (impact phase) of the attacker's kick

		Posterior counterattack	Action that begins after the opponent's attack (during the descending phase, or when attacker's leg touches the ground) with a scoring purpose. Athletes kick at the same time. This action (sometimes) includes a previous backward displacement to dodge the opponent's attack
Type of technique	Changeable	Linear	The kicking leg is directed toward the front of the opponent's body with a pushing motion in an attempt to kick the opponent with the sole of the foot
		Circular	The kicking leg is directed toward the opponent's side, with a circular movement in an attempt to kick the opponent with the instep
		With previous spin	Action performed with a previous rotation, at least 180° from the initial position, before kicking the opponent
Height target	Changeable	Head	Kick to permitted areas of the head
		Trunk	Kick to permitted areas of the trunk
Laterality	Changeable	Right	Kick performed with the right leg
		Left	Kick performed with the left leg
Kicking leg	Changeable	Forward	Kick performed with the leg closest to the opponent
		Backward	Kick performed with leg furthest from the opponent
Guard	Changeable	Open	The front leg of each opponent differs (i.e., one of them has the left leg advanced and the other the right leg)
		Close	The front leg of both opponents is the same (e.g., both opponents have the left leg advanced)
Score	Changeable	0 pts	Action does not impact on the permitted areas, or impacts in these areas but not with enough force to score
		1 pts	Score obtained by a valid action performed to the trunk with a linear or circular technique
		2 pts	Score obtained by a valid action performed to the trunk using a spin beforehand
		3 pts	Score obtained by a valid action performed to the head with a linear or circular technique
		4 pts	Score obtained by a valid action performed to the head using a spin beforehand
		Kyong-go	Warning from the referee
		Gam-jeon	Penalty by the referee

Note. In mixed criteria, the category selected is fixed during one part of the bout, although it may be changed throughout. In fixed criteria, the selected category does not change during the bout; in this case, these categories were pre-established by the WTF. In changeable criteria, a behaviour could fit either category (that it belongs to) during the bout. Each action to permitted areas with enough strength to displace them is considered to be a valid action (WTF, 2012).

Data collection

Three observers were trained in both category recognition and codification, maintaining the homogeneity of inter and intrasessional and consistency of the observation periods (Anguera et al., 2011). Three observers were responsible for viewing and codify the videos using a consensual concordance. According to the observational methodology (López-López et al., 2015), three observers are more appropriate than two to discuss disagreements and chose the proper code which has to be assigned to the behaviour viewed (consensual concordance strategy). Codification was carried out using HOISAN 1.5.6 software (Hernández-Mendo, López-López, Castellano, Morales, & Pastrana, 2012). The reliability of the registration codification made by the observers was confirmed by: (a) a qualitative approach through consensual agreement (Usabiaga et al., 2013), consisting of previous agreement between the observers as to which code should be assigned to each behaviour; (b) calculating inter and intraobserver agreement (e.g., Cohen's kappa and correlations); and (c) a study of generalizability (Blanco-Villaseñor et al., 2014; Morales-Sánchez, Pérez-López, Morquecho-Sánchez, & Hernández-Mendo, 2016).

Statistical analysis

Linear, normal and homoscedastic distribution of the data was checked using component analysis of variance by maximum likelihood (GLM) and least squares (LS) procedures (Hemmerle & Hartley, 1973; Searle, Casella & McCulloch, 1992) using SAS software (Schlotzhauer & Littell, 1997) in order to check whether the residual errors were equal. Inter and intraobserver concordance was carried out using a calculation of Kendall's tau-b (τ) and Cohen's kappa (κ) indexes using HOISAN software (Hernández-Mendo et al., 2012). The analysis of generalizability was carried out using SAGT software (Hernández-Mendo, Blanco-Villaseñor, Pastrana, Morales-Sánchez, & Ramos-Pérez, 2016), testing the model o × p × c ('o' being observers, 'p' participants, and 'c' categories). Three measurement plans (CO/P, CP/O and PO/C) were extracted from this model, and are related to the three objectives of this study: (a) the measurement plan CO/P allows for the determination of whether the observed behaviour (six bouts of the Olympics) can be extrapolated to the total population (male taekwondo Olympians); (b) the measurement plan CP/O allows for the assessment of inter and intraobserver agreement, determining whether the observers and the codifications of the study are reliable; and (c) the measurement plan PO/C allows for testing of the good adjustment of the categories (E/ME). In each of these models, the generalizability study provides two coefficients: one is the generalizability coefficient (the G relative coefficient, e^2) and the reliability coefficient (the G absolute coefficient, Φ) (Brennan, 2003).

These coefficients range from 0 to 1.0; a value of 1.0 represents perfect reliability, while values ranging between 0.70 and 0.90 are optimal (Lafave & Butterwick, 2014). Values lower than 0.70 indicate unreliable data, which cannot be generalized. The implication of coefficients depend on each model: when the plan CO/P tends to one indicates that the sample used is sufficient to generalise the data to the population. If the coefficient tends to zero it means that athletes behaviour is different between bouts, and therefore it is needed to analyse a higher number of combats. When the plan CP/O tends to one indicates that observers recognize and codify the same actions in the same way, that is, codifications are homogeneous. When the coefficient tends to zero indicates a poor agreement between observers, being necessary a longer training period of them. Finally, when the plan PO/C tends to one indicates a homogeneous categories and overleaping of categories, being necessary to review the categories included in the observational tool, while it tends to zero indicates the good adjustment of categories, which are heterogeneous. To use GT effectively, it is necessary to determine the components of the variance, which generate errors in design, and the relative weights of these components (Blanco-Villaseñor et al., 2014).

RESULTS

The results of the generalizability analysis for each criterion of the observational tool are shown in Table 2. Inter and intraobserver rates calculated for Kendall's tau-b and Cohen's kappa are shown in Table 3.

TABLE 2
Variance component and G coefficients (e^2 and Φ) of the criteria in the model O × P × C for six combats analysed.

	Criterion 1 Round		Criterion 2 Weight Category		Criterion 3 Event		Criterion 4 Colour protector		Criterion 5 Tactical action		Criterion 6 Type of technique		Criterion 7 Height target		Criterion 8 Laterality		Criterion 9 Kicking leg		Criterion 10 Guard		Criterion 11 Score	
	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS	df
Model	82240.7	51	246314.5	18	45407.7	149	80020.7	49	107541.3	34	107541.3	34	107541.3	34	107541.3	34	107541.3	34	107541.3	34	125390.0	59
C	306.8	2	3037.5	1	1013.7	8	3280.5	8	4654.4	8	4654.4	8	4654.4	8	4654.4	8	4654.4	8	4654.4	8	3019.9	8
P	3096.9	8	1006.3	6	7170.4	8	12410.1	2	22685.5	1	22685.5	1	22685.5	1	22685.5	1	22685.5	1	22685.5	1	53409.8	5
O	595.8	2	448.3	2	157.1	2	481.1	2	528.3	2	528.3	2	528.3	2	528.3	2	528.3	2	528.3	2	144.4	2
C×P	464.5	16	0.0	0	7064.5	63	7822.4	16	3111.6	8	3111.6	8	3111.6	8	3111.6	8	3111.6	8	3111.6	8	4331.9	17
C×O	50.1	4	0.0	0	108.2	6	371.9	6	257.8	6	257.8	6	257.8	6	257.8	6	257.8	6	257.8	6	151.0	6
P×O	364.0	6	178.6	4	612.2	16	83.4	4	271.5	2	271.5	2	271.5	2	271.5	2	271.5	2	271.5	2	706.8	8
O×P×C	95.3	12	0.0	0	715.4	45	626.9	10	509.8	6	509.8	6	509.8	6	509.8	6	509.8	6	509.8	6	237.7	12
Error	1510.3	9	582.5	2	1229.3	24	3159.3	8	2613.7	6	2613.7	6	2613.7	6	2613.7	6	2613.7	6	2613.7	6	1923.0	9
Total	83751.0	60	246897.0	20	46637.0	173	83180.0	57	110155.0	40	110155.0	40	110155.0	40	110155.0	40	110155.0	40	110155.0	40	127313.0	68
CO/P	$e2 = 0.8$ Φ = 0.7		$e2 = 0.9$ Φ = 0.9		$e2 = 0.7$ Φ = 0.5		$e2 = 0.9$ Φ = 0.6		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9	
CP/O	$e2 = 0.9$ Φ = 0.8		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9		$e2 = 0.9$ Φ = 0.9	
PO/C	$e2 = 0.9$ Φ = 0.9		$e2 = 1.0$ Φ = 0.4		$e2 = 0.7$ Φ = 0.7		$e2 = 1.0$ Φ = 0.9		$e2 = 0.1$ Φ = 0.1		$e2 = 0.1$ Φ = 0.1		$e2 = 0.3$ Φ = 0.1		$e2 = 0.3$ Φ = 0.1		$e2 = 0.3$ Φ = 0.1		$e2 = 0.3$ Φ = 0.1		$e2 = 0.6$ Φ = 0.1	

Note. SS= Sum of squares (ANOVA type III), df = degrees of freedom, G coefficient relative = e^2, G coefficient absolute = Φ. Letters: C, P & O correspond to the facets of categories, participant and observers respectively. G models: PO/C shows if categories fulfil the condition of being E/ME. CP/O explains the observers' agreement according to the participants and categories of the observational tool. CO/P shows whether the number of participants selected is sufficient to generalize the behaviour to the universe

TABLE 3
Inter and intraobserver agreement.

	Interobserver						Intraobserver
	Bout 1			**Bout 2**			
	1&2	1&3	2&3	1&2	1&3	2&3	
τ	0.95	0.97	0.95	0.99	0.99	0.99	0.95
κ Round	0.94	0.98	0.95	0.96	0.97	0.97	0.91
κ Weight category	0.94	0.98	0.95	0.96	0.97	0.97	0.91
κ Event	0.94	0.98	0.95	0.96	0.97	0.97	0.91
κ Colour protector	0.94	0.95	0.91	0.95	0.94	0.95	0.89
κ Tactical action	0.81	0.88	0.73	0.92	0.93	0.94	0.70
κ Type of technique	0.90	0.83	0.83	0.92	0.93	0.94	0.82
κ Height target	0.94	0.89	0.87	0.94	0.93	0.97	0.80
κ Laterality	0.92	0.89	0.87	0.88	0.84	0.92	0.84
κ Kicking leg	0.92	0.91	0.87	0.94	0.90	0.94	0.85
κ Guard	0.84	0.88	0.77	0.91	0.91	0.86	0.78
κ Score	0.90	0.88	0.80	0.95	0.92	0.95	0.85

Note. τ means Kendall's tau-b (calculated by using every criterion in the same pool). κ means Cohen's kappa (calculated in each criterion). To determine interobserver agreement, the indexes (τ and κ) were obtained by observers (comparison in pairs) for the analysis of the same bout (files 1&2, 1&3, and 2&3 provide the interobserver agreement, each number corresponding with that of the observer) while the analysis of the same bout twice was pre-established, to give a measure of the intraobserver agreement.

DISCUSSION AND CONCLUSIONS

The objective of this study were: (a) determine whether the number of bouts selected, six taekwondo bouts from those in the Olympic Games in London in 2012, allows the generalization of the Olympians' behaviour to the corresponding population (plan CO/P); (b) determine the inter and intraobserver agreement and reliability (plan CP/O); and (c) assess the accuracy of the E/ME categories present in each of the criteria of the observational tool for obtaining reliable data (plan PO/C).

With regard to the sample size needed in order to generalize the data obtained to a population with similar characteristics to the sample (i.e., taekwondo Olympic male competitors), the results (plan CO/P) showed G coefficients (e^2 and Φ) with values greater than 0.7; these values are considered sufficiently reliable to generalize the analysed behaviour to all taekwondo athletes who participated in the Olympic Games (75 bouts with 68 competitors). Similarly, previous studies in the field of sports (Blanco-Villaseñor et al., 2000; Usabiaga et al., 2013) showed generalizable data with small samples of between one and 12 matches, assuming at least 5%–10% of the total sample (Balmaseda, 2013; Blanco-Villaseñor et al., 2000; Usabiaga et al., 2013). Specifically, a previous study in combat sports (i.e., Box) was used as a preliminary study with six bouts (two of each weight category) to estimate the

technical-tactical behaviour needed to generalize boxers' behaviour. They had to analyse 30 bouts (10 of each weight category) due to the variable behaviour among combats (Balmaseda, 2013). Similar qualities of results were found in the current study with the analysis of six combats in taekwondo athletes. This ability to analyse only a sample of the entire population underlines the power of generalizability studies, which allows researchers to extrapolate from data when it is not possible to analyse the entire sample, thus saving time and costs.

With regard to observer reliability, the Cohen's kappa and Kendall's tau-b values (τ and κ) were calculated to give a measure of interobserver (comparison in pairs for the analysis of the same bout) and intraobserver agreement (for the analysis of the same bout twice). In this sense, the results of using observer agreement (consensual agreement in coding behaviours) showed highly reliable data, since values were above 0.7 for Cohen's kappa (Cohen, 1960) and Kendall's tau-b indexes (Landis & Koch, 1977; Banerjee et al., 1999), confirm the optimum observers and data reliability (τ = 0.95 to 0.99; κ = 0.70 to 0.98) (Table 3). These results support the results of the generalizability analysis plan CP/O, which allows the determination of interobserver reliability, which also showed highly reliable observations (0.90) in all criteria (Table 2). The fact that values tend to 1.0 (perfect agreement) indicates that observers are able to recognize the sporting behaviours and do not generally make errors of omission (not encoding an action when it has occurred) or commission (encoding an action that has not occurred) during the codification process (Blanco, 1989). Likewise, intraobserver agreement (observations made by the same observer over two different time periods) is also optimal (above 0.7). These results also affirm that rigor has been preserved in coding behaviours throughout the process of data collection. These results are in line with previous studies in sports field (Balmaseda, 2013; Pradas, Floría, González-Jurado, Carrasco, & Bataller, 2012; Tarragó et al., 2017; Usabiaga et al., 2013) showing similar values above 0.7 in their results. In this sense, the authors of the current work agree with previous studies (Medina & Delgado, 1999; McKenzie, 2010) on the importance of the prior training of observers, both in terms of codifying observations with specific software and recognizing behaviours during taekwondo bouts, and of monitoring the results to ensure the reliability of the observations. This is important to ensure the necessary objectivity and reliability of data collected using observational methodology.

Finally, regarding the accuracy and validity of the categories of the observation tool (plan PO/C), the values obtained with the model tend to 1.0 for the fixed and mixed criteria (those defined by the sport's regulations) and tend to 0 for the changeable criteria (those concerning to the technical/tactical behaviour of competitors) (Table 4). The tendency towards zero shows that categories are exhaustive and mutually exclusive (E/ME) (that is,

heterogeneous) for each criterion (Blanco-Villaseñor et al., 2014), allowing the categories used to analyse the technical/tactical behaviour of taekwondo athletes to be stated, and this can explain the behaviour in a reliable way complying with the requirements of the observational methodology (Anguera et al., 2007). More implications of E/ME categories are the observational manual created by researchers, in which behaviours and its codification are explained, is adequate and there is no confusion in which category codify each behaviour, that is, no overleaping of categories (Anguera et al., 2007). As for the fixed criteria (those which are unchanged throughout the observational session) and mixed criteria (those which are unchanged during a part of the observational session and changed in another part), these are determined by the rules of the World Federation Taekwondo (WTF, 2012), such as the number of rounds within each bout, weight categories, the event within the championship and the colour of the protector. For these criteria, the results of the present study showed homogeneous categories (G values tended to 1.0). These results seem to be congruent, since they cover all possible categories of each criterion (Shavelson & Webb, 1991); these must define the competition system and cannot be modified. Similar results were found in previous studies in combat sports (Balmaseda, 2013), highlighting the E/ME condition needed for the observational tools designed to reflect the real behaviour of competitors. Other study also stated the necessary bibliography revision and experts' consultation to ensure the validity of the observational, and the quality of the data (Tarragó et al., 2017). Therefore, the results of the present study allow researchers to assume a finer adjustment of the observational tool (a mixed system of format field and categorical system), due to the optimal values of the generalizability coefficients which represent a stronger indication of the observational tool's reliability (Lafave & Butterwick, 2014), in addition to ensure the validity of the observational tool for this sport.

In regard to the limitations of this study, it should be noted that changes to the regulations have recently occurred (WTF, 2015) in which kicks to the truck with a spin beforehand now score three points, that is, an additional point to the scoring used in the last Olympic Games in London (2012). In addition, it should be noted that differences emerged between genders (Menescardi et al., 2012), then, the results of the current study only could be generalized to male population. Further research is therefore necessary in this context to include studies with the new scoring rules and in female competitors, in order to check whether the technical/tactical patterns have changed between the two championships and between genders.

Therefore it is possible to conclude that the present design of the generalizability study (estimation plans CO/P, CP/O and PO/C) showed highly reliable (e^2) and generalizable values (Φ) (close to 1.0) for the sample size, the

inter and intraobserver agreement, and reliability of the criteria (and categories) of the observational tool. These results allow researchers to generalize the data and observations made, and indicate that the research design is accurate enough to pass quality control tests (Blanco-Villaseñor et al., 2014). Moreover, the observational tool is a valid instrument which could be used in future research to analyse behaviours within taekwondo, and particularly within the Olympic Games. In addition, this study provides the basis for developing future observational studies of technical/tactical level, which in turn could be applied in practice sessions, where taekwondo coaches and athletes depend on video analysis to improve their technical/tactical skills.

ACKNOWLEDGEMENTS

Cristina Menescardi was supported by the Conselleria d'Educació, Cultura i Esport of the Generalitat Valenciana under grant VALi+d [ACIF/2014/268]. Antonio Hernández-Mendo is supported by the Ministerio de Economía y Competitividad [Secretaría de Estado de Investigación, Desarrollo e Innovación] under grant [DEP2012-32124] and the Junta de Andalucia [Conserjería de Innovación, Ciencia y Empresa] under grant [SEJ 444].

REFERENCES

Anguera, M. T., Blanco, A., Hernández-Mendo, A., & Losada, J. L. (2011). Diseños Observacionales: Ajuste y aplicación en psicología del deporte [Observational designs: Their suitability and application in sports psychology]. *Cuadernos de Psicología del Deporte, 11*(2), 63-76.

Anguera, M. T., Magnussson, M. S., & Jonsson, G. K. (2007). Instrumentos no estándar: Planteamiento, desarrollo y posibilidades [Non-standard instruments: Approach, development and possibilities]. *Avances en Medición, 5*, 63-82.

Balmaseda, M. (2013). *Escuela cubana de boxeo: análisis de las acciones técnico-tácticas (SOBOX)* [Cuban box school: Technical-tactical actions analysis (SOBOX)]. Sevilla: Wanceulen.

Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures*. Canadian Journal of Statistics, 27*(1), 3-23.

Blanco-Villaseñor, A. (1989). Fiabilidad y generalización de la observación conductual [Reliability and generalization of the behavioral conduct]. *Anuario de Psicología, 43*(4), 5-32.

Blanco, A., Castellano, J., & Hernández-Mendo, A. (2000). Generalizabilidad de las observaciones de la acción del juego en el fútbol [Observations of the football game action: A Generalizability study]. *Psicothema, 12*(Supl. 2), 81-86.

Blanco-Villaseñor, A., Castellano, J., Hernández-Mendo, A., Sánchez-López, C. R., & Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra [Application of the generalizability theory in sport to study the validity, reliability and estimation of samples]. *Revista de Psicología del Deporte, 23*(1), 131-137.

Brennan, R. L. (2003). *Coefficients and Indexes in Generalizability Theory***.** Iowa, IA: University of Iowa.

Casolino, E., Lupo, C., Cortis, C., Chiodo, S., Minganti, C., Capranica, L., & Tessitore, A. (2012). Technical and tactical analysis of youth taekwondo performance. *Journal of Strength and Conditioning Research, 26*, 1489-1495.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological measurement, 20*, 37-46.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York, NY: John Wiley and Sons.

Del Vecchio, F. B., Franchini, E., Del Vecchio, A. H. M., & Pieter, W. (2011). Energy absorbed by electronic body protectors from kicks in a taekwondo competition. *Biology of Sport, 28*(1), 75-78.

Falco, C., Estevan, I., Álvarez, O., Morales-Sánchez, V., & Hernández-Mendo, A. (2014). Tactical analysis of the winners' and non-winners' performances in a Taekwondo University Championship. *International Journal of Sport Science & Coaching, 9,* 1407-1416*.*

Falco, C., Landeo, R., Menescardi, C., Bermejo, J. L., & Estevan, I. (2012). Match analysis in a university taekwondo championship. *Advances in Physical Education, 2*(1), 28-31.

Hemmerle, W., & Hartley, H. (1973). Computing maximum likelihood estimates for the mixed AOV Model using the w-transformation. *Technometrics, 15*, 819–831.

Hernández-Mendo, A., Blanco-Villaseñor, A., Pastrana, J. L., Morales-Sánchez, V., & Ramos-Pérez, F. J. (2016). SAGT: Aplicación informática para análisis de generalizabilidad [SAGT: software to analyze the Generalizability Theory]. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte, 11*(1), 77-89.

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales, V., & Pastrana, J. L. (2012). HOISAN 1.2: Programa informático para uso en Metodología Observacional [IT Program for use in methodology observational]. *Cuadernos de Psicología del Deporte, 12*, 55-78.

Kazemi, M., Casella, C., & Perri, G. (2009). 2004 Olympic tae kwon do athlete profile. *The Journal of the Canadian Chiropractic Association, 53*(2), 144-152.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lafave M. R., & Butterwick, D. J. (2014). Generalizability theory study of athletic taping using the technical skill assessment instrument. *Journal of Athletic Training, 49*(3), 368–372.

Martínez, R. (1995). Teoría de la generalizabilidad (I): Conceptos básicos. Estudios G.: Modelos de efectos aleatorios [Generalizability theory (I): Basic concepts. G studies: Random effect models]. In R. Martínez. *Psicometría. Teoría de los tests psicológicos y educativos.* Madrid: Síntesis.

McKenzie, T. L. (2010). Observing physical activity and its contexts. *Research Quarterly for Exercise and Sport, 81*(2), 113–122.

Medina, J., & Delgado, M. A. (1999). Metodología de entrenamiento de observadores para investigaciones sobre E.F. y deporte en las que se utilice como método la observación [Methodology of observer's training for investigations about PE and sports in which methodology observational was used]. *Motricidad, 5,* 69-86.

Menescardi, C., Bermejo, J. L., Herrero, C., Estevan, I., Landeo, R., & Falco, C. (2012). Diferencias técnico-tácticas en taekwondistas universitarios según sexo y categoría de competición [Technical-tactical differences among university level taekwondo competitors by gender and weight division]. *Revista de Artes Marciales Asiáticas, 7*(2), 1-11.

Menescardi, C., López-López, J. A., Falco, C., Hernández-Mendo, A., & Estevan, I. (2015). Tactical aspects of a national university taekwondo championship in relation to round and match outcome**.** *Journal of Strength and Conditioning Research, 29*, 466-471.

Mérida, R., Serrano, A., & Tabernero, C. (2015). Diseño y validación de un cuestionario para la evaluación de la autoestima en la infancia [Design and validation of a questionnaire for children's self-esteem assessment]. *Revista de Investigación Educativa, 33*(1), 149-162.

Morales-Sánchez, V., Pérez-López, R., Morquecho-Sánchez, R., & Hernández-Mendo, A. (2016). Generalizabilidad y gestión deportiva [Generalizability and Sports Management]. *Cuadernos de Psicología del Deporte, 16*(1), 161-170.

Pradas, F., Floría, P., González-Jurado, J. A., Carrasco, L., & Bataller, V. (2012). Developement of an observational tool for single table tennis analysis. *Journal of Sport and Health Research, 4*(3), 255- 268.

Santos, V., Franchini, E., & Lima-Silva, A. (2011). Relationship between attack and skipping in taekwondo contests. *Journal of Strength and Conditioning Research, 25*(6), 1743-1751.

Schlotzhauer, S. D., & Littell, R. C. (1997). *SAS system for elementary statistical analysis* (2nd ed.). Cary, NC: SAS Institute.

Searle, S., Casella, G., & McCulloch, C. (1992). *Variance components***.** New York, NY: John Wiley and Sons.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling Variability of Performance Assessments. *Journal of Educational Measurement, 30*(3), 215-232.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Thousand Oaks, CA: Sage.

Tarragó, R., Iglesias, X., Lapresa, D., Anguera, M. T., Ruiz-Sanchis, L., & Arana, J. (2017). Análisis de las relaciones diacrónicas en los comportamientos de éxito y fracaso de campeones del mundo de esgrima utilizando tres técnicas complementarias [Analysis of diachronic relationships in successful and unsuccessful behaviors by world fencing champions using three complementary techniques]. *Anales de Psicología, 33*(3), 471-485.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1978). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.* Washington, DC.: Department of Health, Education and Welfare (DHEW Publication No. OS 78-0012). Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm

Tornello, F., Capranica, L., Minganti, C., Chiodo, S., Cibellis, G., & Tessitore, A. (2014). Technical-Tactical Analysis of Youth Olympic Taekwondo Combat. *Journal of Strength and Conditioning Research, 28*, 1151-1157.

Usabiaga, O., Castellano, J., Blanco-Villaseñor, A., & Casamichana, D. (2013). La Teoría de la Generalizabilidad en las primeras fases del método observacional aplicado en el ámbito de la iniciación deportiva: calidad del dato y estimación de la muestra [Generalizability theory in the early stages of the observational method applied in the field of sports initiation: Data quality and sample size estimation]. *Revista de Psicología del Deporte, 22*(1), 103-109.

Wickel, E. R., & Welk, G. J. (2010). Applying generalizability theory to estimate habitual activity levels. *Medicine & Science in Sports & Exercise, 42*(8), 1528-1534.

World Taekwondo Federation (2012). *Competition rules.* Retrieved from http://www.wtf.org/wtf_eng/site/rules/competition.html

World Taekwondo Federation (2015). *Competition rules.* Retrieved from http://www.wtf.org/wtf_eng/site/rules/competition.html.

Yiannakos, A., & Armatas, V. (2006). Evaluation of the goal scoring patterns in European Championship in Portugal 2004. *International Journal of Performance Analysis in Sport, 6*(1), 178-188.