Korean Journal of Sport Science
2020, Vol. 31. No. 4, 830-839
https://doi.org/10.24985/kjss.2020.31.4.830

체육과학연구

Analysis of inter-rater agreement of latin american and modern dance sport

Seung-hun Lee & Mi-sun Kim*

Korea Institute of Sport Science

[Purpose] The purpose of this study is to identify the consistency and correlation of referee evaluation according to the judging characteristics in the preliminary and semi-final of the Latin event of dance sports, thereby deriving the problem of referee reliability and suggesting alternatives for improvement. **[Methods]** The method of study is Based on the performance data of 54 amateur Latin dances match organized by the Korea Dance Sports Federation for a total of three years 11,850(preliminary rounds & semi-final). Based on the Kappa statistics and the degree of agreement(p_a), the difference in the group of examination characteristics was derived and the correlation between the five Latin events was analyzed. **[Results]** As a result of through this study, the consistency of the dance sports referee and the characteristics of the judging in the event were confirmed, and the number of judges tended to be higher when the number of judges was seven, the more the number of subjects was, and the highest level of agreement was more than 70 percent of the judges. In addition, the higher the concordance of each of the five detailed items, the higher the correlation tendency. **[Conclusion]** Differences in visual aspects between dance sports judges and the difference in the judges' ratings due to the revision of the rules, the decrease in the number of competitors participating in the competition, the number of people to be eliminated in each round, and the proportion of judges with experience in the competition are different, and these differences affect the judges and show up in the scores. The Latin dance sports events based on objectivity and reliability to improve the correct standards of judges to find the same raters, work will be required and an assessment element. With a systematic way in and to carry out the review curriculum and educational development is considered necessary.

Key words: dance sports, amateur latin events, judging consistency, Kappa coefficient, referee reliability

서 론

스포츠에서는 선수와 지도자뿐만 아니라 예리한 판단력이 요구되는 심판의 역할이 경기에서 중요한 요소로 인식되고 있다(Kim, 2016; Shin & Kim, 2017). 특히 기술력뿐만 아니라 예술성을 내포하고 있는 스포츠 종목(리듬체조, 피겨스케이팅, 다이빙, 아티스틱 스위밍, 댄스스포츠)의 경우 심판의 판정이 승패 또는 순위를 결정하기 때문에 수행 기록 및 측정 기록을 통해 평가되는 다른 스포츠 종목(역도, 육상, 사격 등)보다 심판의 주관적인 판단에 대한 의존도가 높다(Kim, Woo & Kim, 2011; Oh & Lee, 2002; Cho, Eom & Han, 2017; Premelč, Vučković, James & Leskošek, 2019; Kim, & Jung, 2008.; Cho & Choi, 2015).

예술성을 내포하고 있는 스포츠 중 하나인 댄스스포츠(Dancesport)는 최소 7명 이상의 심판들이 각기 다른 루틴(routine) 수행하고 있는 다수의 커플에 대한 정보를 취합하여, 주관적인 평가를 통해 승패를 결정하는 경기이다(Kim, 2019). 심판들은 경험적 지식에 기초하여 기술

논문 투고일 : 2020. 09. 10.
논문 수정일 : 2020. 10. 21.
게재 확정일 : 2020. 11. 19.
* 교신저자 : 김미선(imskyelove@gmail.com).

적 요소와 미적 요소를 빠르고 정확하게 파악할 책임이 따르는데, 각 심판이 각 선수를 평가하는데 걸리는 시간은 대략 예선경기는 3.75초(1분30초/24커플), 준결승경기는 7.50초(1분30초/12커플)소요되며, 결승경기는 대략 7.71초 시간 동안 각 선수의 평가가 이루어진다(Hur, 2006; Kim, 2019). 이렇듯 짧은 시간 동안 다수의 커플의 특징, 행동, 표현, 기술 등을 관찰하여 정확한 판정을 도출해 내는 것은 쉬운 일이 아니다(Radler, 1998). 따라서, 댄스스포츠 심판들은 선수나 지도자 경험이 무엇보다 중요하며, 공정성 및 전문성을 바탕으로 객관적인 판단을 할 수 있는 능력이 중요하다(Kim, Lim & Ha, 2018; Cho, 2007).

이러한 중요성에 따라 지도자 및 심판 교육과 심판 배정은 대한민국댄스스포츠연맹(KFD : Korean Federation of Dancesport)에서 담당하여, 정기교육(10시간), 의무교육(60시간), 보수교육(스케이팅 시스템 관련 교육) 등을 수행하고 있다(대한민국댄스스포츠연맹, 2020). 그리고, 매년 1회 이상 정기교육을 받아야 심판 배정을 받을 수 있으며, 심판은 국제댄스스포츠연맹(IDSF : International Dance Sports Federation)과 WD(World Dance)/DSC(Dance Sport Council)의 심사 기준을 따른다. 라틴댄스의 경우 가장 중요한 요소인 파워(Lee & Hur, 2006)를 비롯해, 자세(posture), 라인(line), 홀드(hold), 균형(poise), 박자(timing), 조화(togetherness), 음악성(musicality), 표현력(expression), 발과 다리 동작(foot and leg), 리드와 팔로우(lead and follow), 플로어 사용(floor craft), 의상, 안무 등에 대한 요소들을 종합적으로 평가하며, 심판에 따라 중요하다고 생각하는 평가요소가 다를 수 있다(Radler, 1998). 이렇듯 주관적인 평가에 따라 서로 다른 결과를 초래할 수 있어 심판간의 신뢰도는 매우 중요한 부분이라 할 수 있는데, 심판 신뢰도(rater reliability)란 심사자의 평정 결과가 일치하는 정도로 2명 이상의 평가자에 의하여 부여된 점수의 일치정도로 정의된다(Baumgartner et al, 2006).

지금까지 댄스스포츠 심판의 주관적 심사 결과의 신뢰도 검증을 위해 다양한 방법으로 연구했는데, 선행연구를 살펴보면 Rasch 평정척도 모형을 이용하여 일치도 및 심판 판정의 적합도(Kim & Jeon, 2018), 순위상관계수를 사용한 심판 일치도 분석(Jeong & Kim, 2006; Kwon, 2011), 심사 척도 개발(Kim & Jeong, 2008) 등 많은 연구가 이루어지고 있다. 하지만 지금까지 댄스스포츠 심사 일치도 연구는 모두 결승전 순위 결정 시스템에 의한 판정의 일치도를 알아보기 위해 사용되었고, 예선 판정 시스템인 통과 유무를 기록한 범주형 자료에 대한 분석은 미흡한 실정이다. 댄스스포츠 예선은 상대적으로 많은 선수들을 짧은 시간에 평가하는데, 일반적인 판정 방법인 점수 또는 순위체점이 아닌 정해져있는 통과 범주에 맞게 통과 유무를 결정하는 시스템이다. 이에 따라 심판의 주관적 판정에 의해 다음라운드 진출 여부 및 탈락 여부가 경기 평가 중 가장 짧은 시간에 결정되는 만큼 심사의 신뢰도는 매우 중요한 문제라고 할 수 있다.

따라서, 본 연구는 라틴 종목 예선과 준결승에서 심판들의 일관성을 평가하기 위해 라운드, 연도, 대회, 종목, 심판 특성에 따른 심판 평가의 일치도와 상관성을 파악하여 심판 신뢰도의 문제를 도출하고 개선을 위한 대안을 제시하는데 목적이 있다.

연구방법

연구자료

본 연구에서는 대한민국댄스스포츠연맹(KFD)에 연구의 목적, 방법, 절차를 설명한 후 자료 수집을 허가받았으며, 2017년 ~ 2019년 까지 총 3년간의 연맹 주관으로 실시된 아마추어 라틴 종목 심판 판정 자료를 수집하였다. 연맹 웹 사이트(http://www.kfds.or.kr) 에서는 모든 경기의 선수 순위 및 심판 판정 기록을 경기 후 공개하고 있는데, 총 54경기(2017: 21경기, 2018: 15경기, 2019: 17경기), 11,850건(예선 : 8,610건, 준결승: 3,185건)의 사례수를 분석에 활용했다.

댄스스포츠의 실기능력 평가요소

1. 라틴 종목 판정 기준

Radler(1996)는 기본적인 자세(posture), 라인(line), 홀드(hold), 균형(poise), 박자(timing), 조화

(togetherness), 음악성 및 표현력(musicality and expression), 힘(power), 프레젠테이션(presentation), 발과 다리 동작(foot and leg action), 리드와 팔로우(lead and follow), 마루에서의 기능(floor craft), 의상 및 안무(intangibles) 등을 모두 비교 평가해야 한다고 보고하였다. 국내 전국체전 댄스스포츠 경기 대회에서는 자세, 균형성(balance), 협응력(coordination), 동작의 완성도(quality of movement), 음악에 대한 동작의 일치도(movement to music), 파트너 십(partnering skill), 그리고 작품과 표현력(choreography and presentation) 등 7개의 평가요소를 선정하고 있다(김지은, 전유정, 2018). 또한, 2009년 심판시스템이 발표되면서 자세, 밸런스와 협응(posture, balance and coordination), 움직임의 특성(quality of movement), 음악에 맞는 움직임(movement to music), 파트너 관계(partnering), 안무와 표현(choreography and presentation)으로 나누어 평가로 구성한다고 발표하였다. 이후 2013년부터 심판시스템이 향상된 버전으로 구현되면서 평가요소는 총 4가지(기술의 질, 음악에 맞는 움직임, 파트너기술, 안무와 표현)로 축소되었다(김미선, 2019).

2. 심판선정 방법 및 심사지 작성 내용

대한민국댄스스포츠연맹에서는 댄스스포츠 심판을 위한 전문성과 심판 선발과정의 객관성을 높이기 위해 연맹 자격심의위원회의 공정한 심의를 통과하여 심판 급수를 결정하게 된다. 대한민국댄스스포츠연맹의 댄스스포츠 심판원으로 활동하기 위해서는 심판 자격증을 소지하고 있어야 하며, 매년 심판 정기 교육을 실시하여 이에 대한 교육을 받은 심판원들 한에서 1년간 심판으로 활동할 수 있는 자격이 부여된다. 심판 등급은 총 3등급으로 나뉘며, 본 연구에서는 1급, 2급 A, 2급 B, 3급 A, 3급 B로 나누어 진행하였다. 심판 등급별 배정 범위는 2017년 까지는 규정되어 있지 않았으나, 2018년 개정되어 진행되었으며, 2019년 회장배 및 KPDC 대회, 국가대표 및 상비군 선발전에 대한 대회에 대한 배정 범위가 1급에서 3급 A에 속한 심판원으로 배정 범위가 개정되었다. 댄스스포츠 라틴심사는 5종목으로 구성되어 진행되며 종목 구성 및 경기 진행 순서는 삼바(samba), 차차차(chachacha), 룸바(rumba), 파소도블레(paso doble), 자이브(jive)로 진행된다.

경기가 시작되기 전 경기에 배정된 심판들은 각 종목별 심사지 5장을 가지고 경기장의 가장자리에 자유롭게 자리를 잡고 서 있는 상태에서 경기 진행 순서에 맞추어 음악이 나오면 평가를 시작한다. 심사지에는 무작위로 배포한 출전 선수의 번호가 적혀있고, 만약 출전 선수가 많은 경우 조를 나누어 진행하게 되는데 심사지에는 조별로 나누어 출전 선수의 번호가 모두 적혀있다. 조별로 최대 12커플씩 경기장에 나와 음악이 끝날 때까지 진행된다. 예를 들어, 30명의 출전 선수가 3조로 나누어서 한조씩 나와 춤을 추면 조별로 나누어 평가하는 것이 아니라 출전한 30명의 선수 모두를 고려하여 평가하여 다음 라운드에 진출할 선수(약 24명)의 번호 옆에 통과 여부에 대해 체크하는 형식으로 진행된다. 이렇듯 예선 및 준결승은 선수의 다음라운드에 대한 통과유무를 결정하고, 결승은 선수의 우열을 평가하여 순위를 결정하는 형식으로 진행되는데, 본 연구에서는 예선과 준결승에 대한 심사를 토대로 분석하였다.

심판 판정 일치도 평가 지표

1. Kappa 통계량

동일한 측정 대상들에 대해 평가자들의 평가 결과의 일치 정도를 일치도(agreement)라고 정의한다. 카파통계량(kappa statistic)은 측정한 결과가 범주형 자료일 때 일치도의 척도로 자주 쓰인다. 평가자가 둘일 때는 단순 카파통계량(Cohen, 1960) 또는 가중 카파통계량(Cohen, 1968)이 활용되고, 평가자가 세 명 이상일 때는 Fleiss Kappa 통계량을 활용할 수 있다(Fleiss, 1971; Berry and Mielke, 1988). 카파통계량의 계산식은 다음 (1)과 같은데, 평가자들이 대상자들을 우연히 같은 범주로 분류하는 경우가 있으므로, 그 확률(p_e)을 보정한 일치도를 사용한다.

$$k = \frac{p_a - p_e}{1 - p_e} \quad (1)$$

p_a = Ratio of observations on which there is agreement
p_e = Agreement which is expected by chance alone

본 연구에서는 r명의 심사자가 n명의 선수를 q개의 범주(통과 여부)로 평가한다고 가정할 때, 일치도(p_a)를 계산하기 위해서 (2)와 같이 정의했다.

$$p_a = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q}\frac{r_{ij}(r_{ij}-1)}{r(r-1)} \quad (2)$$

r = Number of judges
n = Number of players
q = Evaluation category
r_{ij} = Number of judges i th who rated the competitor in category j th

2. Kappa 통계량의 제한점

Fleiss의 방법은 Scott의 개념(Scott, 1955)을 확장시킨 형태로 평가 결과의 범주는 명목형이며 평가자는 서로 독립임을 가정하며, 실제 결과로 제시되는 값은 가중치가 적용되지 않는 카파계수(exact Kappa coefficient)로 나타난다(Conger, 1980). Fleiss's kappa 결과는 p_a가 같더라도 주변분포의 주변 동질성과 균형성 여부에 따라 k값이 크게 변화하는 문제를 가지고 있는데, 이로 인해 관찰된 일치 비율이 큼 에도 불구하고 k가 많이 낮아질 수 있다는 것이 Fleiss's kappa 의 큰 문제점이 된다(Kim, Song, Nam, & Jung, 2012; Feinstein, & Cicchett, 1990). 실제 일치도 분석에서 카파통계량은 주변 확률 변화에 따라 값이 의존적인 큰 문제점 때문에 이에 관한 많은 연구가 진행되고 있고, 보완하기 위한 몇몇 방법들이 제안되었다(Brennan & Prediger, 1981; Park & Park, 2007; Gwet, 2010).

3. 수정된 Kappa 통계 산출 방법

k값이 주변분포에 민감하게 되면 실제로 일치도가 높아 보이는 자료라도 계산된 κ값은 그렇지 않은 경우가 종종 발생한다. 또한 본 연구 자료는 2가지 범주(통과 유무)를 선택할 수 있는 경우의 수(합격 수)가 정해져 있는 특성상 주변 동질성과 균형성에 의해 우연의 일치도(p_e) 값이 과대평가 되는 문제가 발생한다. 그로 인해 k값은 과소평가 되는 경향을 보 일 수 있기 때문에 본 논문에서는 κ값이 p_e에 민감하지 않도록 보완된 Kappa 통계량인 Gwet(2008a)의 AC1방법을 활용했으며, 다음 (3)과 같이 정의된다.

$$k_{AC1} = \frac{\frac{1}{nr(r-1)}(\sum_{i=1}^{n}\sum_{j=1}^{q}r_{ij}^2 - rn) - \frac{1}{q-1}\sum_{j=1}^{q}p_j(1-p_j)}{1-\frac{1}{q-1}\sum_{j=1}^{q}p_j(1-p_j)} \quad (3)$$

4. Kappa 통계량의 해석 기준

Kappa 통계량의 해석 기준은 가장 일반화되어 있으며, 널리 활용되고 있는(Donner & Eloasziw, 1987) Landis & Koch(1977)의 기준으로 해석하였다. 평가자 간의 일치도를 분석하기 위해 카파일치도(Kappa measurement of agreement)값과 민감도와 특이도의 가능성을 위해 95%수준에서 CI값(신뢰구간, Confidence interval)을 제시하였다. Kappa 통계량의 해석은 0.81~1.00인 경우 '거의 완벽하게 일치'(almost perfect agreement), 0.61~ 0.80인 경우 '높은 수준의 일치'(substantial agreement), 0.41~0.60은'보통수준의 일치'(moderate agreement), 0.21~0.40의 경우'상당한 일치'(fair agreement), 0.01~0.20의 경우 '약간 일치'(slight agreement), 〈 0인 경우 '일치 기회 이하'(less thanchance agreement)로 해석한다(Landis & Koch, 1977; Viera & Garrett, 2005).

자료분석방법

본 연구에서는 수집된 자료의 처리를 위하여 통계스크립트 기반의 무료 패키지인 R 3.5.1 버전을 사용하였다. Kappa 통계량 계산을 위하여 R프로그램 전용 소프트웨어인 irrCAC 패키지를 함께 사용하여 각 변수별 카파계수와 95% 신뢰구간을 제시하였다. 또한 통계처리(data processing in statistics)를 위해 SPSS Statistics 20 프로그램을 이용하여 기술통계 분석을 실시하였고, 일치도(p_a)를 기준으로 5개의 세부종목, 연도, 라운드, 심판수, 선수출신 심판 비율 집단간 차이 검증을 위하여 일원변량분석(one-way ANOVA)을 보수적 사후검정법인 Scheffe 검정과 함께 사용하여 결과를 도출하였다. 또한

Table 1. Comparison of characteristics ANOVA result of judge's agreement

Characteristics	Categories	N	Kappa coefficient(k)		Positive agreement(p_a)			
			k_{ac1}	95% CI	M	SD	F-value	post-hoc (Scheffe)
Amateur event Latin-5	① Samba	2,370	0.398***	0.381~0.415	0.695	0.199	3.231*	①>④⑤②③
	② Cha cha cha	2,370	0.371***	0.354~0.388	0.680	0.197		
	③ Rumba	2,370	0.360***	0.343~0.377	0.676	0.196		
	④ Paso doble	2,370	0.384***	0.366~0.401	0.688	0.201		
	⑤ Jive	2,370	0.376***	0.359~0.393	0.684	0.197		
Year	① 2017	5,610	0.341***	0.321~0.362	0.658	0.192	108.224***	③>②>①
	② 2018	3,500	0.406***	0.38~0.432	0.697	0.201		
	③ 2019	2,740	0.465***	0.435~0.494	0.723	0.199		
Competition	① Round 1	5,800	0.427***	0.406~0.448	0.704	0.202	62.129***	①②>④③
	② Round 2	2,375	0.427***	0.391~0.463	0.687	0.198		
	③ Round 3	490	0.231***	0.18~0.282	0.606	0.161		
	④ Semi-Final	3,185	0.330***	0.306~0.353	0.660	0.190		
Number of judges	① 7 judges	1,005	0.462***	0.403~0.521	0.703	0.223	6.341***	①>④②③
	② 9 judges	3,470	0.394***	0.367~0.422	0.687	0.204		
	③ 11 judges	6,205	0.372***	0.353~0.391	0.678	0.191		
	④ 13 judges	1,170	0.405***	0.36~0.45	0.696	0.191		
Ratio of judges experience as a player	① 0 ~ 30%	1,400	0.286***	0.248~0.325	0.629	0.172	52.543***	④>②③>①
	② 30% ~ 50%	2,155	0.399***	0.365~0.434	0.687	0.195		
	③ 50% ~ 70%	5,165	0.383***	0.362~0.403	0.684	0.197		
	④ 70% ~ 100%	3,130	0.439***	0.41~0.468	0.708	0.208		

*p<.05, ***p<.001

아마추어 라틴 5개 종목별 일치도 간 상관성을 분석하기 위하여 피어슨의 상관분석(Pearson′s correlation)을 통해 분석을 진행하였으며, 모든 통계상 유의수준은 .05로 설정하였다.

연구결과

심사자간 일치도 분석 결과

수집 자료의 특성(세부종목, 연도, 라운드, 심사자 수, 선수출신 심사자 비율)에 따른 심사 위원 일치도(p_a)의 차이를 알아보기 위하여 총 54경기, 11,850건의 연기를 바탕으로 일원변량분석을 실시하였다. 분석 결과 모든 부분에서 통계적으로 유의한 차이를 나타냈다(p<.05, p<.001). 〈Table. 1〉은 수집 자료의 특성별로 구분된 집단 간의 심사 위원 일치도에 대한 평균과 표준편차, 일원분산분석 검정 결과를 카파계수와 함께 제시한 것이다. 종목의 경우 쌈바(69.5%, k=0.398, 95% CI=0.381~0.415)일치율이 가장 높게 나타났으며, 파소도블레(68.8%, k=0.384), 자이브(68.4%, k=0.376), 차차차(68%, k=0.371), 룸바(69.5%, k=0.360)일치율이 같은 집단으로 구분되며, 통계적으로 유의미한 차이가 나타났다(F=3.231, p<.05).

연도의 경우 2019년(72.3%, k=0.465), 2018년(69.7%, k=0.406), 2017년(65.8%, k=0.341)순서로 해가 거듭될 수 록 일치도가 올라가는 양상을 보이며, 통계적으로 유의미한 차이가 나타났다(F=108.224, p<.001).

라운드별 일치도의 차이를 살펴보면 1라운드(70.4%, k=0.427), 2라운드(68.7%, k=0.427)가 같은 집단으로 구분되며 상대적으로 높은 일치도를 보였으며, 3라운드(60.6%, k=0.231)와 준결승(66.0%, k=0.330)이 같은 집단으로 구분되며 상대적으로 낮은 일치도를 보이며, 통계적으로 유의미한 차이가 나타났다(F=62.129, p<.001).

심사자 수에 따른 차이를 살펴보면 7심(70.3%, k=0.462)일 때 가장 높은 일치도를 나타냈으며, 13심(69.6%, k=0.405), 9심(68.7%, k=0.394), 11심(67.8%, k=0.372)이 같은 집단으로 구분되며. 통계적으로 유의미한 차이가 나타났다(F=6.341, p<.001).

심사자들 중 선수 출신의 비율에 따른 차이를 살펴보면 70% 이상(70.8%, k=0.439)일 때 가장 높은 일치도를 나타냈으며, 30%~50%(68.7%, k=0.399), 50%~70%(68.4%, k=0.383)는 같은 집단으로 구분되었고, 30%이하(62.9%, k=0.286)일 때의 일치도가 가장 낮게 나타나 통계적으로 유의미한 차이가 나타났다(F=52.543, p<.001).

세부 종목별 일치도 상관 분석

댄스스포츠 라틴 종목 5개의 세부종목(삼바, 차차차, 룸바, 파소도블레, 자이브) 사이의 상관분석 결과는 〈Table. 2〉에 나타난 바와 같다.

5개의 세부종목 간에는 모두 높은 수준의 정적 상관관계를 보였다(p<.05). 특히 파소도블레와 자이브의 심사 일치도간 상관관계가 가장 높게 나타났으며(r=0.672, p<.05), 파소도블레와 삼바의 심사 일치도간 상관관계도 높게 나타났다(r=0.668, p<.05). 반면 룸바와 차차차의 심사 일치도간 상관관계는 가장 낮게 나타났으며(r=0.632, p<.05), 룸바와 자이브의 심사 일치도간 상관도 낮게 나타났다(r=0.637, p<.05).

Table 2. Correlation analysis results according to agreement of amateur latin 5

	a	b	c	d	e
a	1				
b	.650*	1			
c	.642*	.632*	1		
d	.668*	.655*	.642*	1	
e	.651*	.641*	.637*	.672*	1

*p<.05, a: Samba , b: Cha cha cha, c: Rumba, d: Pasodoble, e: Jive

논의

본 연구는 댄스스포츠 라틴 종목 심사자 간의 일치도 경향과 세부 종목별 심사 일치도의 상관관계를 확인하기 위한 목적으로 실시되었다. 본 연구에서 도출한 결과에 대한 논의는 다음과 같다.

첫째, 분석 결과를 세부종목별로 종합하여 Kappa 계수를 통해 알아본 댄스스포츠 심판 일치도 경향은 5종목의 일치도 계수 모두 61% 보다 높은 일치를 보였다. 그중에서도 다른 종목에 비해 삼바가 69.5%로 가장 높게 나타났으며, 다음으로 파소도블레(68.8%), 자이브(68.4%), 차차차(68.0%), 룸바(67.6%) 순으로 나타났다. 특히 세부종목별 일치도가 높을수록 상관성도 높은 경향을 보였다. 라틴 댄스는 경기장을 직사각형으로 봤을 때 바깥쪽 라인과 평행하게 시계 반대 방향으로 춤을 진행하는 LOD(line of dance)인 삼바와 파소도블레, 춤의 진행 방향이 자유로운 차차차, 룸바, 자이브로 나눌 수 있다. 심판은 경기가 시작되기 전 경기장의 가장자리에 자리를 잡고 서 있는 상태에서 경기가 시작되면 자유롭게 움직이며 심사를 진행하게 된다. 이때 LOD 종목의 경우 선수마다 다른 루틴의 경로로 심판의 눈앞으로 춤을 추며 시계 반대 방향으로 지나가게 된다. 그로 인해 심판은 선수들의 움직임을 거시적인 시각으로 바라보며 평가를 진행할 수밖에 없었을 것으로 보여 진다. 반대로 룸바의 경우에는 Kim & Jung, 2006 의 연구와 같이 25-27bars/min으로 가장 느린 음악에 맞추어 추는 종목이기 때문에 미시적인 시각으로 바라보며 보다 세밀한 부분까지 평가가 가능했던 것으로 보여 지며, 이는 차차차

도 비슷할 것으로 판단된다. 정리하자면 LOD와 LOD가 아닌 종목의 특이성으로 인해 평가 시 바라보는 시각적인 부분에 차이가 심사 간의 평가점수에 영향을 준 것으로 보여 진다. 그리고 흥미로운 것은 LOD가 아닌 종목 중 자이브 종목이 일치도가 높았다는 것이다. 이러한 결과는 자이브가 경기순서에 있어 가장 마지막에 진행되는 종목이며, 빠른 템포(42-44bpm)에 맞추어 계속해서 뛰는 스텝 위주로 구성되어 있어 체력적인 요소가 두드러지게 차이 나는 경우가 많아 심사 시 심판의 평가요소에 대한 우선순위에 있어 심판간의 의견 차이가 크지 않아 심사 간의 평가점수 일치도가 높았던 것으로 판단된다. 이처럼 다음라운드에 대한 통과유무를 결정하는 예선 및 준결승 심사의 경우 종목에 따라 우선하는 평가요소가 있을 것으로 생각되며, 향후 이 부분에 대한 추가적인 연구가 필요할 것으로 판단된다.

둘째, 연도별로는 연도가 지날수록 심사자 간 일치도가 향상되는 경향을 보이는 것으로 나타났다. 연도별 달라진 규정을 살펴보면, 2017년은 연맹에 소속되어 있는 심판들을 연맹 측에서 배정한다는 규정만 있을 뿐 배정되는 심판에 대한 등급 범위 규정이 정해져 있지 않은 실정이었다. 그러나 2018년을 기점으로 심판 등급별 배정 범위에 대한 규정이 개정되었고, 2019년에는 대회 규모가 큰 회장배, KPDC대회, 국가대표 및 상비군선발전, 그리고 공인대회(아마추어)의 심판 등급 배정 범위 등급이 높아졌다. 이런 규정 개정에 의해 심사자의 심판 등급의 차이가 심사 일치도에 영향을 준 것으로 보인다. 또한, 2017년은 평균 25.57±11.03커플이 출전하였지만, 2018연에는 24.40± 10.23커플, 2019년은 18.50±5.18커플로 점차 대회에 참가하는 선수가 줄어드는 경향을 보였고, 이러한 영향도 배제할 수 없을 것으로 보인다.

셋째, 댄스스포츠는 체조, 아이스댄싱 등과 같은 종목과 다르게 다른 피겨를 수행하는 여러 팀에 대한 평가를 동시에 진행하여 평가하는 형식으로 이루어진다. 예를 들어, 1라운드에 56팀이 출전을 하게 되면 2라운드에 진출하는 선수는 36팀으로 20팀을 탈락시켜야 하며, 대회에 출전한 팀이 적어 1라운드에서 바로 준결승으로 진출을 한다고 하더라도 6~10팀을 탈락시켜야 한다. 따라서 짧은 시간 동안 심판은 다수의 팀을 다음 라운드 진출 여부 및 탈락 여부를 결정해야 한다. 본 연구결과를 살펴보면, 1라운드와 2라운드가 3라운드와 준결승보다 상대적으로 높은 일치도를 보였는데 이는 출전한 팀이 가장 많은 1라운드와 2라운드에서 선수들을 평가하는데 있어 많은 요소를 보고 평가하기 보다는 가장 눈에 띄는 요소를 통해 빠르게 평가하기 때문으로 보여 진다. Radler (1998)는 기본적인 요소(자세, 선, 홀드, 균형, 타이밍, 조화 등)를 보고 그밖에 요소(음악성, 표현력, 힘, 발과 다리 동작, 리드와 팔로우, 마루에서의 기능, 의상, 안무 등)를 비교평가해야 한다고 보고하였고, Lee & Hur (2006)는 선수들을 심사하는데 있어 첫 번째로 고려하는 요소로 음악성과 표현성(musicality & expression), 타이밍(timing)을 뽑았다. 이처럼 탈락시켜야하는 인원수가 많은 경우(1라운드와 2라운드)와 탈락시켜야하는 인원수가 적고, 비슷한 실력을 가진 선수가 많은 경우(3라운드와 준결승)의 평가에 있어 고려해야하는 평가요소가 다를 수 있을 것으로 예상되며, 이에 의한 차이가 심사 간의 평가점수에 영향을 미쳤을 것으로 보인다. 그러나 심사자 간 동일한 평가요소를 통해 평가했다고 보기는 어렵기 때문에 라운드별로 고려해야하는 요소에 관련된 후속연구가 필요하다.

넷째, 본 연구결과에서 심사자 수가 7심일 때 일치도가 가장 높았고, 그다음이 13심일 때였다. 2004년도까지만 해도 세계 댄스스포츠 챔피언쉽의 경우 심사자 수를 최소 11명으로 구성하며, 유럽 댄스스포츠 챔피언쉽의 경우에도 최소 9명의 심판으로 심사위원을 구성하였다. 또한, 국내 대회에서는 대체적으로 이보다 적은 심판으로 구성한다고 하였다(Choi et al., 2004). 그러나 본 연구에서 사용된 자료를 살펴보면 2017년도부터 2019년도까지 7심으로 진행된 대회는 7개뿐 이였고, 이것 또한 2017년에 집중되어 있었다. 9심은 16개, 11심은 28개, 13심은 연도별로 1개씩 총 3개의 대회가 있었다. 말하자면 대부분 9~11심으로 진행되었고, 7심으로 이루어진 대회의 경우 참가선수의 수가 9심~13심에 비해 현저히 적었던 것을 알 수 있었다. 따라서 이에 대한 결과는 두 집단의 사례 수에 비해 현저히 낮기 때문에 비교하는데 제한점이 따를 것으로 보인다.

다섯째, 댄스스포츠 심판은 타 스포츠 활동을 했던 사람이 심사하기에는 무리가 있으며(Kim, Lim & Ha, 2018), 대한민국댄스스포츠연맹 측에서도 심판자격증을

취득하는데 있어 응시자격에 경기인 출신에 대한 공적사항으로 심판자격증 승급에 대한 혜택이 이루어지고 있다(Korean Federation of Dancesport, 2020). 또한, 심판의 선수경험이 선수들을 평가하는데 직접적인 영향을 미치는지에 대한 의문은 많은 연구자들이 제시한 바 있다(Cho, 2009; Jeong & Kim, 2006; Choi, 2006). 본 연구결과를 살펴보면, 심판배정에 있어 선수경험이 있는 심판의 비율이 높아질수록 심사 일치도가 올라갔으며, 선수경험이 있는 심판의 비율이 70%이상일 때, 70.8%, 카파계수 0.439(보통수준의 일치)로 상대적으로 높은 심사 일치도를 보였다. 여기서 흥미로운 것은 심판배정에 있어 선수경험이 있는 심판의 비율이 2017년 61.62±16.49%, 2018년 60.90±15.11%, 2019년 54.89±13.71%로 점점 떨어진다는 것이다. 그럼에도 불구하고 연도가 지날수록 심사자 간의 일치도가 향상된다는 것은 기존 연구에서 심판의 선수경험 여부가 심판평가에 영향을 미치지 않는다는 결과와 일치하는 것으로 판단할 수 있으나(Lee & Hur, 2006; Jeong & Kim, 2006), 연도별로 보았을 때 모두 70%이상이 아니었기 때문에 이에 대한 해석은 주의가 필요할 것으로 판단된다.

결론 및 제언

본 연구는 2017년부터 2019년까지 대한민국댄스스포츠연맹에서 제공하는 공식 경기 기록 중 예선 및 준결승 심판 판정 결과를 바탕으로 라운드, 연도, 대회, 종목, 심판 특성에 따른 심판 평가의 일치도와 상관성을 파악하여 심판 신뢰도의 문제를 도출하고 개선을 위한 대안을 제시하는데 목적이 있다. 본 연구의 결과를 토대로 내린 결론은 다음과 같다.

본 연구결과를 종합적으로 살펴보면, 댄스스포츠 심사자 간의 시각적인 부분에 대한 차이와 규정 개정으로 인한 심판 등급 차이, 대회 참가 선수의 감소, 라운드별 탈락시켜야 하는 인원수, 선수 경험이 있는 심판의 비율 등의 차이로 다르게 나타나게 되고 이러한 차이로 인해 심사자에게 영향을 미쳐 심사점수에 나타난 것으로 보인다. 이 연구를 통해 댄스스포츠 심판 일치도와 종목에 대한 심사 특성을 확인하였는데, 단일 대회에 대한 자료가 아닌 수년간에 걸친 심판 기록 자료를 바탕으로 한 연구 결과라는 점에서 괄목할 만한 점이다. 이를 바탕으로 댄스스포츠 라틴 종목 심판들의 객관성과 신뢰도 향상을 위해 정확한 기준을 찾는 작업이 필요하며, 심사자들이 동일한 평가요소를 가지고 심사를 진행할 수 있도록 교육과정에 있어 체계적인 방법 및 교육 프로그램 개발이 필요할 것으로 생각된다.

참고문헌

Baumgartner, T. A., Jackson A. S., Mahar, T. M., & Rowe, D. A. (2006). Measurement for evaluation in physical education and exercise science (8th). Dubuque, IA : WCB McGraw-Hill.

Berry, K. J., & Mielke, P. W. (1988). A generalization of Cohen's kappa. *Educational and Psychological Measurement, 48*, 921-933.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement, 41*(3), 687-699.

Cho, E. H., & Choi, Y. L. (2015). Analysis of error sources in results of evaluation of difficulty(D) and execution(E) by judges of rhythmic gymnastics competition. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science 17*(3), 13-22.

Cho, E. H., Eom, H. J., & Han, Y. S. (2017) Analysis of multiple error source in results by judges of artitstic gymnastics women's uneven bars events. *Korean Society For Measurement And Evaluation In Physical Education And Sports Science, 19*(3), 57-69.

Cho, M. H. (2007). *A qualitative analysis on expertise of dance sport judge*. Dongduk Women's University Graduate School.

Cho, M. J. (2009). Influence of Judges' decision cognition of hockey players on psychological states and game. *Korean Journal of Sports Science, 18*(2), 339-348.

Choi, B. I., Won, Y. S., & Kim, S. Y. (2004). Comparative study about dance sport's marking system: To the total system and skating system. *Journal of p.e., sport & leisure studies, 11*(1), 27-40.

Choi, Y. S. (2006). Examining rater errors in dance

performance assessment. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science, 8*(1), 81-93.

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

Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin, 70*(4), 213.

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin, 88*(2), 322.

Donner, A., & Eliasziw, M. (1987). "Sample size requirements for reliability studies." *Stat Med, 6*(4), 441-449.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology, 43*(6), 543-549.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters, Psychological Bulletin, 76, 378-382.

Gwet, K. L. (2008a). Computing inter rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*(1), 29-48.

Gwet, K. L. (2010). Handbook of Inter-Rater Reliability, 2nd edn. Advanced Analytics, LLC. Janson, H. and Olsson, U. (2001). *A measure of agreement for interval or nominal multivariate observations,*

Hur, K. A. (2006). *An analysis on judges' criteria for judgment in dancesports competition in korea.* Master Dissertation, The Graduate School of Kyungnam University, Korea, Seoul.

Jeong, J. O., & Kim, E. J. (2006). The judges` organization and objectivity of a dance sport competition. *Journal of Sport and Leisure Studies 26*(1), 483-495.

Kim, D. G. (2016). A study on rigorous introduction of electronic refereeing system & video-based decision system for fairness of sports. *The Korean Association of Sports & Entertainment Law, 19*(3), 45-62.

Kim, E. J., & Jung, J. O. (2008). The mixed method for development of preliminary dancesport judging scale. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science, 10*(1), 13-29.

Kim, E. J., & Jeong, J. O. (2006). Judges` organization of Dancesport competition using the Skating System. *The Korean Journal of Physical Education, 45*(3), 469-479.

Kim, H. D., Woo, D. I., & Kim, E. J. (2011). Objectivity of international aerobics referee decision. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science. 13*(2), 63-73.

Kim, J. E., & Jeon, Y. J. (2018). Ranking determination of dancesport players: application of rasch rating scale model. *The Journal of Korean Dance, 36*(4), 45-67.

Kim, M. S. (2019). *Analysis of judgement system for dancesport by visual search.* Ph. D. Dissertation, Kookmin University, Korea, Seoul.

Kim, M. S., Lim, Y. H., & Ha, H. S. (2018). Trust in dancesport competition judges and moral philosophy of kant. *Journal of the Korean Society for the Philosophy of Sport, Dance & Martial Arts, 26*(3), 111-120.

Kim, M. S., Song, K. J., Nam, C. M., & Jung, I. (2012). A study on comparison of generalized kappa statistics in agreement analysis. *Korean Journal of Applied Statistics, 25*(5), 719-31.

Korean Federation of Dancesport. (2020). Regulations for judging management. *KFD Judgment Committee. http://www.kfd.or.kr/*

Kwon, S. R. (2011). *The comparison of a rating agreem ent and a grading judgement among international and domestic judges in dance sport.* Master Dissertation, Kyung Hee University, Korea, seoul.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics, 1*, 159-174.

Lee, C. H., & Hur, K. A. (2006). A analysis on judges' criteria for judgement in dance sports competition. *Korea Sport Research, 17*(5), 253-264.

Oh, S. H., & Lee, B. Y. (2002). Introduction to objectivity and its usage. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science 4*(1), 1-12.

Park, M. H., & Park, Y. G. (2007). A new measure of agreement to resolve the two paradoxes of Cohen's Kappa. *The Korean Statistical Society, 20*(1).117-132

Premelč, J., Vučković, G., James, N., & Leskošek, B. (2019). Reliability of judging in dancesport. *Frontiers in psychology, 10*(1), 1001.

Radler, D. (1998). How a dance competition is judged.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public opinion quarterly,* 321-325.

Shin, J. E., & Kim, Y. j. (2017). Identifying the key elements of judging in synchronized swimming: for the purpose of fair decision. *Asian Journal of Physical Education and Sport Science, 5*(1), 41-55.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med, 37*(5), 360-363.

댄스스포츠 라틴 종목 심사 특성에 따른 일치도 분석

이승훈, 김미선
한국스포츠정책과학원 분석연구원

〔목적〕 본 연구는 댄스스포츠 라틴 종목 예선과 준결승에서 심사 특성에 따른 심판 평가의 일치도와 상관성을 파악하여 심판 신뢰도의 문제를 도출하고 개선을 위한 대안을 제시하는데 목적이 있다. **〔방법〕** 2017~2019 총 3년간 대한민국댄스스포츠연맹 주관 아마추어 라틴댄스 54경기 11,850건(예선, 준결승)의 연기 자료를 바탕으로 Kappa 통계량과 일치도(p_a)를 기준으로 심사 특성 집단 차이를 도출하고, 라틴 5개 종목별 일치도 간 상관성을 분석했다. **〔결과〕** 본 연구를 통해 댄스스포츠 심판 일치도와 종목에 대한 심사 특성을 확인하였으며, 연도가 거듭될수록, 심판 수는 7심일 때, 대상 선수가 많을수록 일치도가 높은 경향을 보였으며, 심사자들 중 선수 출신의 비율이 70% 이상일 때 가장 높은 일치도를 보였다. 또한 5개의 세부종목별 일치도가 높을수록 상관성도 높은 경향을 확인하였다. **〔결론〕** 결론적으로 댄스스포츠 심사자 간의 시각적인 부분에 대한 차이와 규정 개정으로 인한 심판 등급차이, 대회 참가 선수의 감소, 라운드별 탈락시켜야 하는 인원수, 선수 경험이 있는 심판의 비율 등의 차이로 다르게 나타나게 되고 이러한 차이로 인해 심사자에게 영향을 미쳐 심사점수에 나타난 것으로 보인다. 이를 바탕으로 댄스스포츠 라틴 종목 심판들의 객관성과 신뢰도 향상을 위해 정확한 기준을 찾는 작업이 필요하며, 심사자들이 동일한 평가요소를 가지고 심사를 진행할 수 있도록 교육과정에 있어 체계적인 방법 및 교육 프로그램 개발이 필요할 것으로 생각된다.

주요어: 댄스스포츠, 아마추어 라틴 종목, 심사 일치도, 카파계수, 심판 신뢰도