

Analysis of reliability and error sources of judges' rating in rhythmic gymnastics

Tae-Koo Lee, Joo-Hyun Kwak, Hee-Won Yang, & Han-Joo Lee*

Yonsei University

[Purpose] The purpose of this study was (1) to analyze judges' evaluation on rhythmic gymnastics performance by applying generalizability theory and (2) to suggest recommendations to improve judges' rating. **[Methods]** Data were 34 players' scores from Senior Part at 29th KGA President's Cup National Rhythmic Gymnastics Championship in Korea. Difficulty and execution scores in ball, clubs, hoop and ribbon event were analyzed. Analysis models containing components of area and reputation rank were designed and multivariate generalizability theory were used for analysis. **[Results]** The G-study results showed (1) that the error source about players has more significant impact to evaluation than other error sources in analysis model containing components of only player and judge, (2) that the error source about players has more significant impact to evaluation than other error sources in analysis model adding components of area, but the error source about area has more significant impact to evaluation of clubs event than other error sources, (3) that the error source about players has more significant impact to evaluation than other error sources in analysis model adding components of reputation rank, but the error source about reputation rank has more significant impact to evaluation of hoop event than other error sources in analysis model adding components of area. The D-study results showed generalizability coefficient was stable in analysis model without components of area and reputation rank, but generalizability coefficient in analysis model containing components of area and reputation rank not stable in some event. **[Conclusion]** Recommendations for improving judging were discussed.

Key words: Generalizability Theory, Rhythmic Gymnastics, Judges' rating, Reliability, Error Source, Judging bias

서론

리듬체조와 같은 예술적 스포츠는 다른 경기와 다르다. 일반적으로 스포츠 경기는 시간이나 높이, 거리 또는 득점으로 경기수행 결과를 나타내지만, 리듬체조의 경우, 심판이 인식하는 예술성과 같은 요소들이 심판의

주관적인 판단에 영향을 받으며, 경기 결과에 영향을 준다.

리듬체조 경기결과와 심판판정은 난도(Difficulty)와 실시(Execution)의 합산으로 이루어진다. 난도 구성은 선수의 신체난도, 댄스스텝, 회전과 던지기를 동반한 다이나믹 요소 등이며, 실시 구성은 예술적, 기술적 구성을 포함하는 선수의 신체표현 움직임, 기초기술 결여, 수구낙하 등을 평가한다. 구체적으로 난도(D)의 채점방법은 선수가 경기 전 제출한 난도표(Paper)를 확인하며 연기 순서대로 올바르게 수행하는지를 평가한다. 반면 실시(E) 평가는 10점 만점을 기준으로 감점한 점수를

논문 투고일 : 2017. 04. 21.

논문 수정일 : 2017. 05. 22.

게재 확정일 : 2017. 06. 09.

* 교신저자 : 이한주(hlee@yonsei.ac.kr).

제한 나머지 점수를 계산하여 실시점수의 최종점을 산출하게 된다. 실시(E)는 예술성, 기술적 구성을 포함하기 때문에 심판의 주관적 오류가 작용할 수 있다. 예를 들어, Leskošek et al.(2012)은 2011 챔피언십 유럽피안 남자체조 경기에서 심판판정에 대해 조사하였는데, 연구결과 심판과 같은 국적의 선수에게 실시심판(E)에서 높은 점수를 주는 것으로 나타났다.

리듬체조 경기 심판들의 평가는 인위적인 오류를 유발할 수 있기 때문에 체조협회는 교육과 강습을 통해 심판의 주관적 판정 오류를 최소화하기 위한 다양한 시도를 하고 있다. 예를 들면 미국 리듬체조연맹은 규정집에 시합 전에 심판들을 교육하고, 시합 중 특정 팀이나 선수에 대한 개인적 편견을 피하고 객관적인 판정을 하도록 가이드라인을 제시하고 있다. 그럼에도 불구하고 심판들의 판정 오류 시비뿐만 아니라 심판 선발관련 부정 문제도 발생(New York Times, 2013.07.10.; Sportschosun, 2011,10.11)하고 있어서, 이는 심판들의 권위와 판정 신뢰도를 떨어뜨리는 요인으로 작용하고 있다.

심판 판정 오류는 비의도적(random) 오류와 의도적(systematic) 오류의 두 가지로 구분할 수 있다(Leskošek et al., 2012). 예를 들어, 리듬체조 경기에서 나타나는 비의도적 오류는 주로 난도점수에서 나타나는 심판들의 오차를 말한다. 난도표에 작성된 표기와 선수의 수행점수간의 계산착오 같은 경우이다. 이러한 오류는 심판들의 경험부족 또는 점수규정의 잘못된 이해 또는 실수가 원인이 될 수 있다. 의도적 오류는 선수의 실시점수에서 나타나는데 심판들의 점수가 실제보다 높거나 낮은 경우를 말한다. 예를 들어, 심판이 부여하는 예술점수는 주관적이기 때문에 선수들의 수행과 관련 없는 심판들의 개인적인 편견(bias)에 기인할 수 있다.

심판판정에 영향을 미치는 심판들의 편견에 대한 선행 연구들은 대략 네 가지로 구분할 수 있다(Auweele et al., 2004). 첫째는 애국심에 관한 편견이다(Ansorge & Scheer, 1988; Leskošek et al., 2012; Ste-Marie, 1996). 국제경기에서 심판들이 자국선수들에게 편향적으로 유리한 점수를 부여한다는 것이다. Popovic(2000)는 시드니 올림픽 리듬체조 대회에서 국제심판들의 판정 패턴을 분석하였다. 연구결과 심판들이 자신들과 동일한 국적 즉, 자국선수들에게는 높은 점수를 부여하는 반면

타국 선수들에게는 낮은 점수를 주는 경향이 있었다고 보고하였다. 둘째는 후광효과(halo effect) 관련 편견이다(Borman, 1975). 하나의 이벤트를 훌륭하게 수행한 선수에 대하여 심판은 인지된 기대감으로 인해 다른 이벤트에도 일반화하여 후한 점수를 준다는 것이다. 이와 같은 후광효과를 Ste-Marie & Valiquette(1996)는 사전에 기억된 수행결과가 이후에도 자연스럽게 이어지는 것이라고 하였다. 셋째는 선수들의 명성에 기인한 편견이다(Findlay & Ste-Marie, 2004). 심판들이 선수들의 명성 또는 세계 랭킹 등에 따라 높거나 낮은 점수를 주는 경향을 말한다. 마지막으로 동조효과(conformity effect)에 따른 편견이다(Sheer et al., 1983). 심판들이 심판석에 있는 동료 심판들과 유사한 점수를 부여하는 경향이 있다는 것이다.

심판들의 편견은 이러한 개인적인 편향성 뿐 아니라 사회적 관계에 의해서도 나타난다. Moon & Jung(2005)은 심판과 지도자와의 관계, 동료심판과의 사회적 관계, 심판과 선수와의 관계가 경기판정에 영향을 미친다고 보고하였다. 또한 Oh & Kim(2015)의 연구에 의하면 체조 선수들은 심판판정이 지역적 또는 학연지연의 관계에 따라 차이가 있다고 믿고 있었다. 요약하면 경기 구성원들 간의 관계 즉, 심판이 알고 있거나 혹은 친분이 있는 선수, 지도자, 동료심판과의 관계에 따라 심판판정이 달라질 수 있다.

이처럼 심판들의 판정 오류는 심판의 개인적인 편향성과 사회적 관계 등 다양한 원인에 기인함을 보고하고 있다. 따라서 심판 판정과 관련한 연구에서 심판 판정에 대한 선수들의 인식이나 심리적 및 운동 수행 측면에서의 영향을 탐색하는 연구뿐만 아니라, 이제는 심판 판정결과 자체에 초점을 두고, 판정 오류에 영향을 주는 원인이 타당인가에 대하여 실증적 연구가 필요한 시점이다(Thelwell et al., 2013).

최근 체육측정평가 연구에서는 심판판정 양상에 대한 연구를 보다 객관적으로 제시하기 위해 일반화가능도 이론(Generalizability theory)을 적용하고 있다. 일반화가능도 이론은 고전검사이론이 단일 오차원(sources of error)만을 고려하는 것과 달리, 측정(평가)상황에서 연구자가 분석하고자 하는 요인들을 포함하는 측정 모형을 구성하고, 이에 따라 구성하는 오차요인을 동시에 분

석한다. 그리고 모형에서 오차점수에 기여하는 다양한 요인들의 영향력을 분해하여 그 값을 수량적으로 구분·제시할 수 있는 측정방법이다(Kim et al., 2010). 또한 측정(관찰) 점수의 오차요인들에 미치는 관계성이 무엇인지 파악하며 올바른 신뢰도를 확보하기 위해 효율적인 측정설계를 제공해 주는 장점이 있다(Kim, 2001).

심판관정에 일반화가능도 이론을 적용한 Lee et al. (2016) 연구는 피겨스케이팅 평가에서 심판의 채점결과를 분석하였다. 연구결과 평가에 가장 많은 영향을 미치는 요인으로는 선수들의 수행으로 나타났으며, 일반화가능도 계수 .9이상으로 높은 신뢰도를 확보하였다. 이러한 결과는 심판이 선수들의 수행능력에 맞춰 객관적인 평가가 이루어졌다는 것을 의미한다. 또한 Cho & Choi(2015)연구에서는 리듬체조대회 심판 8명을 대상으로 후프, 리본종목에서 난도와 실시영역에서 채점한 점수로 일반화가능도 이론을 적용하여 분석하였다. 연구결과 선수요소의 분산성분이 상대적으로 가장 크게 나타났다. 이렇게 두 연구 결과에서 선수요소의 분산성분이 다른 요소들에 비해 가장 크다는 것은 피겨스케이팅과 리듬체조 심판들이 각각 선수들의 수행능력에 맞춰 객관적인 평가를 수행했음을 의미한다.

하지만 일반화가능도 분석에서 선수요소의 분산분석 값의 비율이 가장 클 경우, 한 가지 확인해야 할 점은 잔차의 분산분석 성분 비율이다. 일반적으로 고전검사이론에서 잔차(residual)는 종속변수와 독립변수의 관계를 밝히는 측정모형에서 추정된 종속변수 값과 실제 관찰된 종속변수 값과의 차이를 말하는데, 측정모형에서 추정의 정밀도의 정도를 파악하는데 유용하다(Korea Institute for Curriculum and Evaluation, 2004). 일반화가능도 분석에서 제시되는 잔차는 존재는 하지만, 그 영향력의 출처를 밝히지 못하는 영역이다. 연구자가 일반화가능도 분석을 위해 설계한 분석모형 안에서, 설정한 분석 요인들 외에 다른 요인들의 영향력이 클수록 잔차의 분산성분 값은 커진다. 결국 판정을 하는 평가자가 고도로 훈련되어, 판정을 선수들의 수행능력대로 한다면, 일반화가능도 분석에서 선수 요소의 분산 성분의 비율이 커질 것이고, 자동적으로 잔차의 분산성분 값은 낮을 것으로 추론할 수 있다.

이러한 추론의 정당성은 일반화가능도 이론을 적용한

평가 상황 연구에서 잔차의 분산성분 비율을 살펴보면 알 수 있는데, 학교 교수학습 상황에서 이루어진 동료평가 상황 연구(Kang & Lee, 2006; Lee & Yang, 2016)는 각각 40.5%와 21.4%였으며, 언어 자격증 취득을 위한 평가 상황 연구(Kim et al., 2010)는 최대 9.92%, 엘리트 스포츠 전문 심판 판정 연구(Cho & Choi, 2015; Lee et al., 2016)는 각각 최대 24.3%와 3.5%였다. 이러한 결과들은 Cho & Choi(2015)의 연구 결과를 제외하면, 판정자의 판정 전문성이 확보되면 될수록 잔차의 비율이 낮아진다고 볼 수 있다. 그러나 Cho & Choi(2015)연구에서 훈련된 심판들의 판정 자료를 분석했음에도 불구하고, 후프 실시(E) 영역의 선수와 심판의 상호작용 분산인 잔차의 분산성분(24.3%)이 체육 수업 동료평가를 연구한 Lee & Yang(2016)의 분산 성분 수준(21.4%)보다 높게 나타난 결과는 이해하기 어렵다. 왜냐하면 이러한 결과는 전문심판들이 체육수업 동료 학생평가자들보다 평가 대상자의 실력 외에 다른 요인들에 더 영향을 받고 있다는 것을 의미하기 때문이다. 그렇다면 Cho & Choi(2015) 연구에서 분산성분 24.3%가 나타난 원인은 무엇인가? 이는 이들의 연구에서 적용한 측정모형이 하나의 원인일 수 있다. 심판의 역할은 기본적으로 선수의 경기력에 대한 평가가 집중적으로 이루어져야 하지만 결과와 같이 24.3%의 알 수 없는 요인은 판정의 공정성을 떨어뜨릴 수밖에 없다. 이러한 영역이 미지의 영역으로 남겨진다는 것은 분석을 위해 적용한 측정모형의 추정의 정밀도가 높지 않을 수 있다는 것을 의미한다.

다시 말하면, 리듬체조 심판들은 선수의 경기력 외에 다른 요소들에도 영향을 받고 있기 때문에, 기존의 측정모형에서는 잔차의 비중이 상대적으로 높다고 할 수 있다. 따라서 추정의 정밀도를 높이기 위해서는 다른 요인들을 추가한 새로운 측정모형이 필요하다. 결과적으로 측정 점수의 오차요인들에 미치는 요소들 간의 관계성을 보다 구체적으로 확인할 수 있게 되는 것이다.

따라서 이 연구의 목적은 심판관정 관련 선행 연구들을 기초로 새로운 측정 요소들을 포함한 측정모형을 적용하여 리듬체조 연기 영역 평가에서 나타난 심판 평가의 신뢰도와 오차요인을 조사하는 것이다. 구체적인 연구문제는 다음과 같다. Lee et al.(2016), Cho &

Choi(2015) 연구의 엘리트 경기에서 일반화가능도 이론을 적용하여 심판 판정의 신뢰도를 조사한 것과 같이 첫째, 리듬체조 연기 영역 채점상황에서 일반화가능도에 영향을 주는 요소들의 상대적 크기는 어느 정도인가를 분석할 것이다. 둘째, Ansorge & Scheer(1988), Moon & Jung(2005) 그리고 Oh & Kim(2015) 등이 주장했던 특정 지역에 따라 심판판정이 관계가 있는지를 조사할 것이다. 이들의 주장은 근접한 지역에 활동하기 때문에 자주만나는 관계, 즉 친밀한 사회적관계가 형성되며, 정보를 교환하고, 이렇게 이루어진 연결고리가 심판판정을 편향적으로 만든다는 것이다. 또한 Kwak et al.(2016)이 주장했던 것처럼 수도권지역에 집중되어 있는 지도자들과 다르게 지방에서 훈련을 하는 지도자들은 새로운 루틴이나 수시로 바뀌게 되는 경기관련내용에 대한 정보인지가 다소 느리다. 선수 지도를 위한 정보교환이 불리하다는 말이다. 때문에 심판판정이 지역별로 오차를 나타내는지 조사할 것이다. 셋째, 선행연구는 후광효과가 심판판정의 오류원인이라고 지적하였다(Borman, 1975). 이에 근거하여 이 연구는 지도자의 후광효과가 심판판정과 관계가 있는지를 조사할 것이다. 유명 지도자는 지속적으로 뛰어난 선수를 배출한다. 따라서 심판들은 유명한 지도자들의 선수에 대하여 편향된 기대를 형성하게 되고, 편향된 판정을 할 수 있다. 누구의 제자이기 때문에 뛰어난 선수일 것이란 기대감을 갖게 되고, 그러한 기대에 부응하는 후한 점수를 부여할 수 있다는 것이다. 따라서 이 연구는 심판판정이 유명 지도자에게 배우는 선수와 그렇지 않은 선수에 따라 차이가 있는지에 대해 분석할 것이다.

연구방법

연구 자료

연구자들은 본 연구에서 세 가지의 연구 자료를 수집하였다. 첫 번째 자료는 심판들의 채점결과 자료이다. 본 연구에서는 제29회 회장배전국리듬체조 대회에 참가한 선수들 중 시니어 선수들의 후프(N=31), 볼(N=31),

곤봉(N=31), 리본 종목(N=32)을 채점한 심판들의 채점결과를 사용하였다. 대회에 참가한 고등부 선수들은 현재 국가대표 상비군, 국가대표 선수들도 포함되어 있다. 선수들은 소속 학교를 더미변수로 하여 서울(N=14), 경기도(N=7), 그 외 지역(N=10)의 세 그룹으로 나누어 분석하였다. 이는 대회에 참가한 선수들의 지역 중 서울과 경기 출신이 가장 많았으며, 그 다음으로 부산 인천 충북 등이었기 때문이다. 서울과 경기도를 제외한 지역들은 소수의 인원으로 적절한 분석을 위해 그 외 지역으로 구분하였다.

두 번째 자료와 세 번째 자료는 심판관련 연구 자료로 Kwak et al.(2016)에서 수집·분석된 자료들을 활용하였다. 먼저 대회에 참여한 심판의 인구학적 정보들을 포함한 두 번째 자료는 심판의 기본적인 정보인 선수경력, 심판 자격증 등급, 지도경력(대표선수, 상비군), 협회관련 관계 등을 물어보는 질문들을 포함하였는데, 모든 심판들은 대한체조협회에 소속되어 있고 3급 이상의 자격을 가지고 있었다(Table 1).

Table 1. Participants

	Characteristics	N
Career	school representative	37
	national representative junior	22
	national representative senior	24
Coaching experience	1 ≤ yrs < 5	19
	5 ≤ yrs < 10	10
	yrs ≤ 10	8
Judge certification level	level 3	18
	level 2	5
	level 1 (international)	4

지도자 명성에 따른 심판 분류를 위해, Kwak et al.(2016)의 연구 결과자료를 활용하였다. Kwak et al.(2016)의 리듬체조 지도자 네트워크 분석결과에서 지도자 또는 선수간의 인지도가 가장 많은 지도자들은 중앙성 값이 높았는데, 이들은 지도 선수의 우승경력, 5명 내외의 국가대표선수 배출의 경력 그리고 전년 개인 국가대표, 청소년 국가대표, 상비군지도자 경력이 있는 지도자들이었다. 반면 중앙성 값이 낮은 하위지도자들

은 일반 선수들 지도 외에 국가대표 급 선수를 배출한 경력이 없었다. 연구자들은 중앙성 값을 기준으로 전체 지도자들의 Reputation Rank(상위25% 8명, 하위25% 8명)를 만들어 구분하였다.

채점 방법

리듬체조 대회에서 심판구성은 규모에 따라 조금씩 차이가 있지만 국내의 전국규모 리듬체조 대회는 보통 4명의 난도심판(D)과 4명의 실시심판(E)으로 구성된다(Table 2). 난도(D)는 신체난도(점프, 피벗, 회전), 댄스스텝, 회전요소, 매스터리 등의 연기에서 선수들이 미리 제출한 양식에 따라 채점하여 가산점을 주게 된다(FIG). 그리고 실시(E)는 예술적 결점, 기술적 결점으로 다시 나뉜다. 예술적 결점은 구성의 통일성, 음악과 동작, 신체표현을 포함하고 있으며, 기술적 결점에는 수구의 움직임, 신체동작의 정확성 등에 대해 감점을 하게 된다. 그 밖의 계시심, 라인심에서도 감점 요소가 있다. 총 점수 배점은 각 난도(최대10점), 실시(최대10점)인데 가장 높은 점수와 낮은 점수를 제외한 후, 나머지 점수들을 합하여 평균한 최종점수를 산출한다.

Table 2. Score Distribution and Calculation

Difficulty(D)	Execution(E)
	Max. 10.00 points
	-Artistic Faults
Max.10.00 points	-Technical Faults
4 judges	4-5 judges
By addition:	5 judges + 2 Reference Judges
average of the 2 middle scores + 2	for World Championships / Olympic Games and the other
Reference judges for World Championships	Multisport Games listed in the technical regulations
/ Olympic Games, and the other	By deduction:
Multisport Games listed in the technical regulations	5 judges: average of the 3 middle scores
	4 judges: average of the 2 middle scores
Final Score: 20.00 points maximum	
By addition: D score of 10.00 points maximum + E score of 10.00 points maximum	

자료 분석 방법

리듬체조 대회 원본 채점지는 대한체조협회에서 난도와 실시심판들이 채점한 총 점수가 기록되어 있는 기록 자료를 받아 사용하였다. 각 심판들의 실명은 명시되어 있지 않고, 난도심판의 경우 D1~D4로 표기되고 실시심판의 경우 E1~E4로 표기되어 있다.

자료 분석을 위해 다변량 일반화가능도 이론에 의한 분석을 실시하였다. 본 연구에서 적용한 설계는 구체적으로 선수의 지역 또는 지도자의 명성에 따른 선수 집단의 구분을 적용하지 않은 $p^* \times r^*$ 설계와, 지역 또는 지도자 명성으로 선수 집단의 구분을 적용한 $(p^* : a^*) \times r^*$, $(p^* : a^*) \times r^*$ 설계로 나누어진다.

첫 번째 설계는 연구대상인 선수를 모든 심판들이 판정한 상황으로, 이를 일반화가능도 분석 모형으로 표현하면 $p^* \times r^*$ 설계에 해당한다. 이 설계는 모든 선수(p)를 모든 심판(r)이 판정하며, 심판은 각각 난도(D)를 판정하는 심판과 실시(E)를 판정하는 심판으로 나누어져 있다. 이 두 판정영역 국면(ν)은 측정 절차를 반복해도 변하지 않는 고정효과(fixed effect)로 정의하며, 고정효과인 국면에 내재된 심판 국면(r)은 열린 원(\circ)으로, 무한 전집에서 표집하는 것으로 가정하는 무선희과인 나머지 국면들은 닫힌 원(\bullet)으로 표시한다(Fig. 1). 아래 그림에서 점선으로 표시된 원이 고정효과로 정의되는 판정영역 국면(ν)이다.

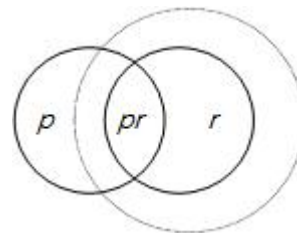


Fig. 1. Venn diagram for $p^* \times r^*$ design

두 번째 설계는 연구대상인 선수를 지역(서울·경기·그 외 지역)에 따라 구분한 상황으로, 이를 일반화가능도 분석 모형으로 표현하면 $(p^* : a^*) \times r^*$ 와 같다. 이 설계는 모든 선수(p)를 모든 심판(r)이 판정하되, 선수가 특정 지역(a)에 속해 있음을 의미한다(Fig. 2).

아래 그림에서 점선으로 표시된 원이 고정효과로 정의되는 판정영역 국면(ν)이다.

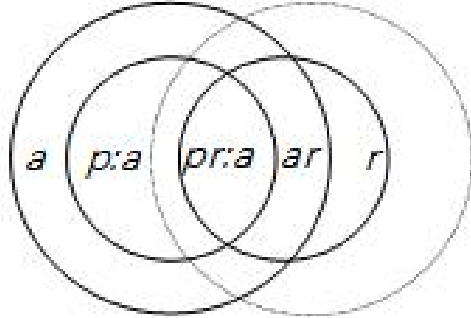


Fig. 2. Venn diagram for $(p^* : a^*) \times r^*$ design

세 번째 설계는 연구대상인 선수를 지도하는 지도자의 Reputation Rank로 구분한 상황으로, 이를 일반화가능도 분석 모형으로 표현하면 $(p^* : l^*) \times r^*$ 와 같다. 이 설계는 모든 선수(p)를 모든 심판(r)이 판정하되, 선수가 지도자의 랭킹(l)에 속해 있음을 의미한다 <Fig. 3>. 아래 그림에서 점선으로 표시된 원이 고정효과로 정의되는 판정영역 국면(ν)이다.

이와 같이 다변량 일반화가능도 이론을 적용하여 G연구(일반화연구)를 통해 분산성분과 공분산성분을 추정하였고, D연구(결정연구)를 통해 일반화가능도 계수를 추정하였다. 다변량 일반화가능도 설계의 분산성분과 공분산성분 추정 및 일반화가능도 계수 추정을 위해서 mGENOVA(Brennan, 2001b) 컴퓨터 프로그램을 사용하였다.

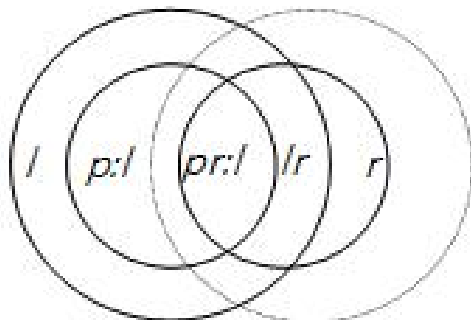


Fig. 3. Venn diagram for $(p^* : l^*) \times r^*$ design

연구 결과

$p^* \times r^*$ 설계에 의한 G연구

$p^* \times r^*$ 설계의 각 분산과 공분산성분에 해당하는 추정치, 해당 분산성분이 전체 분산에서 차지하는 비율, 그리고 판정영역 간 측정오차를 고려한 상관계수는 <Table 4>와 같다. 괄호()는 판정영역별로 분산성분이 전체분산에서 차지하는 비율을 나타낸다. 여기에서 p 는 선수 효과, r 는 심판 효과, 그리고 pr 은 선수와 심판의 상호작용 효과, 즉 잔차를 표시한다.

Table 4. ANOVA for $p^* \times r^*$ design

Event	Effect(α)	Difficulty	Execution
Ball	p	1.3260(84.2)	<1.0000>
		0.8451	0.5348(79.2)
	r	0.0285(1.8)	0.0081(1.2)
	pr	0.2210(14.0)	0.1319(19.5)
	Total	1.5756	0.6749
Clubs	p	1.5588(88.9)	<0.9514>
		0.9825	0.6841(77.4)
	r	0.0392(2.2)	0.0233(2.6)
	pr	0.1561(8.9)	0.1763(20.0)
	Total	1.7541	0.8838
Hoop	p	1.3885(86.4)	<0.9504>
		0.8228	0.5398(81.1)
	r	0.0203(1.3)	0.0000(0.0)
	pr	0.1980(12.3)	0.1257(18.9)
	Total	1.6068	0.6656
Ribbon	p	1.2983(84.2)	<0.9529>
		0.8834	0.6619(86.5)
	r	0.0000(0.0)	0.0000(0.0)
	pr	0.2430(15.8)	0.1034(13.5)
	Total	1.5413	0.7653

분석 결과 난도와 실시영역 두 판정은 유사한 결과를 보였다. 모든 종목이 동일하게 난도영역과 실시영역에서 측정대상인 선수의 분산성분($\hat{\sigma}^2(p)$)이 가장 큰 값을 나타냈다. 꺾은 괄호(<>) 안의 숫자는 판정영역별 측정오차를 고려한 상관계수로, 높은 값을 보였다. 이는 난도영역에서 높은 점수를 받은 선수는 실시영역에서도 높은 점수를 받았음을 의미한다. 본 연구에서는 Brennan(2001a)과 Lee et al.(2015)이 제안한 방법에 따라, 분산성분 추정치의 값이 음의 값으로 산출된 경우 0으로, 측정오차를 고려한 상관계수가 1보다 큰 경우 1.0000으로 표시하였다.

$(p^* : a^*) \times r^\circ$ 설계에 의한 G연구

$(p^* : a^*) \times r^\circ$ 설계의 각 분산과 공분산성분에 해당하는 추정치, 해당 분산성분이 전체 분산에서 차지하는 비율, 그리고 판정영역 간 측정오차를 고려한 상관계수는 <Table 5>와 같다. 괄호(())는 판정영역별로 분산성분이 전체분산에서 차지하는 비율을 나타낸다. *a*는 선수가 속한 지역 효과, *p:a*는 특정 지역에 속한 선수 효과, *r*은 심판 효과, *ar*은 선수가 속한 지역과 심판의 상호작용 효과, 그리고 *pr:a*는 잔차, 즉 앞서 언급한 효과를 제외하고 선수의 점수에 영향을 주는 모든 영향들을 의미한다.

분석 결과 난도와 실시영역은 유사한 결과를 보였으나 종목별로 차이가 나타났다. 곤봉 종목의 경우 지역의 분산성분($\hat{\sigma}^2(a)$)이 가장 높은 값을 보였다(난도 1.6131, 실시 0.7682). 이는 각 판정 영역별 전체 분산성분에서 난도는 70.2%, 실시는 66.7%를 차지하는 값이다. 이는 선수가 어느 지역에 속해 있는지가 선수의 점수에 가장 큰 영향을 주는 요인이 되었다는 것을 의미한다. 곤봉 종목에서 두 번째로 높은 분산성분 값을 보이는 요인은 난도영역에서는 선수($\hat{\sigma}^2(p:a)$), 실시영역에서는 잔차($\hat{\sigma}^2(pr:a)$)였으나 실시영역에서 잔차와 선수요인의 분산성분의 차이는 크지 않은 것으로 나타났다. 곤봉 종목을 제외한 나머지 세 종목에서는 선수의 분산성분($\hat{\sigma}^2(p:a)$)이 가장 높게 나타났다. 이 세 종목의 난도영역과 리본의 실시영역은 선수 요인 다음으로 .

Table 5. ANOVA for $(p^* : a^*) \times r^\circ$ design

Event	Effect(α)	Difficulty	Execution
Ball	a	0.3298(19.5)	<1.0000>
		0.1877	0.1059(14.9)
	p:a	1.1076(65.6)	0.4646(65.2)
		0.7208	
	r	0.0287(1.7)	0.0086(1.2)
	ar	0.0000(0.0)	0.0000(0.0)
	pr:a	0.2213(13.1)	0.1329(18.7)
Total	1.6873	0.7121	
Clubs	a	1.6131(70.2)	<0.9857>
		1.0973	0.7682(66.7)
	p:a	0.4904(21.3)	0.1753(15.2)
		0.2557	
	r	0.0382(1.7)	0.0262(2.3)
	ar	0.0028(0.1)	0.0000(0.0)
	pr:a	0.1543(6.7)	0.1819(15.8)
Total	2.2987	1.1515	
Hoop	a	0.3065(17.8)	<1.0000>
		0.1726	0.0906(13.0)
	p:a	1.1855(68.8)	0.4798(68.6)
		0.7085	
	r	0.0245(1.4)	0.0012(0.2)
	ar	0.0000(0.0)	0.0000(0.0)
	pr:a	0.2063(12.0)	0.1280(18.3)
Total	1.7228	0.6997	
Ribbon	a	0.3498(21.1)	<1.0000>
		0.3361	0.3059(35.2)
	p:a	1.0642(64.2)	0.4572(52.7)
		0.6584	
	r	0.0000(0.0)	0.0000(0.0)
	ar	0.0000(0.0)	0.0000(0.0)
	pr:a	0.2449(14.8)	0.1049(12.1)
Total	1.6588	0.8680	

지역의 분산성분이 높은 값을 보였고, 볼과 후프의 실시 영역은 잔차의 분산성분($\hat{\sigma}^2(pr:a)$)이 더 높게 나타났다. 꺾은 괄호(<>) 안의 판정영역별 측정오차를 고려한 상관계수는 높은 값을 보였다. 즉 선수의 지역을 고려한 경우에도 난도영역에서 높은 점수를 받은 선수는 실시영역에서도 높은 점수를 받았다고 볼 수 있다.

$(p^* : l^*) \times r^*$ 설계에 의한 G연구

$(p^* : l^*) \times r^*$ 설계의 각 분산과 공분산성분에 해당하는 추정치, 해당 분산성분이 전체 분산에서 차지하는 비율, 그리고 판정영역 간 측정오차를 고려한 상관계수는 <Table 6>과 같다. 괄호(())는 판정영역별로 분산성분이 전체분산에서 차지하는 비율을 나타낸다. l 은 선수의 지도자의 명성, $p:l$ 는 특정 지도자 명성에 속한 선수 효과, r 은 심판 효과, al 은 선수의 지도자 명성과 심판의 상호작용 효과, 그리고 $pr:l$ 는 잔차, 즉 앞서 언급한 효과를 제외하고 선수의 점수에 영향을 주는 모든 영향들을 의미한다.

분석 결과 난도영역과 실시영역은 유사한 결과를 보였으나 종목별로 차이가 나타났다. 후프 종목의 경우 특정 지도자 명성의 분산성분($\hat{\sigma}^2(l)$)이 가장 높은 값을 보였다(난도 0.8733, 실시 0.3783). 이는 각 판정영역별 전체 분산성분에서 44% 이상을 차지하는 값이다. 후프 종목을 제외한 나머지 세 종목에서는 특정 지도자 명성에 속한 선수의 분산성분($\hat{\sigma}^2(p:l)$)이 가장 높게 나타났다. 그러나 이 세 종목에서 선수의 분산성분 다음으로 특정 지도자 명성의 분산성분이 높은 값을 보였으며, 특히 이들 세 종목 중에 볼과 곤봉 종목에서 지도자 명성의 분산성분과 선수 분산성분이 전체 분산성분에서 차지하는 비율의 차이가 크지 않게 나타났다. 이는 선수 개인의 영향뿐만 아니라 선수의 지도자의 명성 역시 선수의 점수에 영향을 주는 요인이 될 수 있는 것을 의미한다.

꺾은 괄호(<>) 안의 판정영역별 측정오차를 고려한 상관계수는 높은 값을 보였다. 즉 지도자의 명성을 고려한 경우에도 난도영역에서 높은 점수를 받은 선수는 실시영역에서도 높은 점수를 받았다고 볼 수 있다.

Table 6. ANOVA for $(p^* : l^*) \times r^*$ design

Event	Effect(α)	Difficulty	Execution
Ball	l	0.7698(40.6)	<1.0000>
		0.4764	0.2947(36.7)
	$p:l$	0.8691(45.8)	
		0.5624	0.3599(44.9)
	r	0.0323(1.7)	
			0.0113(1.4)
Clubs	al	0.0000(0.0)	
			0.0000(0.0)
	$pr:l$	0.2265(11.9)	
			0.1365(17.0)
	Total	1.8978	0.8024
Hoop	l	0.8914(42.0)	<1.0000>
		0.6049	0.3960(37.8)
	$p:l$	1.0297(48.5)	
		0.6235	0.4491(42.9)
	r	0.0406(1.9)	
			0.0241(2.3)
Hoop	al	0.0028(0.1)	
			0.0000(0.0)
	$pr:l$	0.1582(7.5)	
			0.1775(17.0)
	Total	2.1227	1.0467
Hoop	l	0.8733(44.5)	<1.0000>
		0.5831	0.3783(46.1)
	$p:l$	0.8702(44.4)	
		0.4767	0.3153(38.4)
	r	0.0068(0.3)	
			0.0000(0.0)
Hoop	al	0.0332(1.7)	
			0.0034(0.4)
	$pr:l$	0.1783(9.1)	
			0.1237(15.1)
	Total	1.9617	0.8207
Ribbon	l	0.3836(22.6)	<0.9790>
		0.3095	0.2605(29.8)
	$p:l$	1.0655(62.7)	
		0.6955	0.5039(57.7)
	r	0.0000(0.0)	
			0.0012(0.1)
Ribbon	al	0.0000(0.0)	
			0.0000(0.0)
	$pr:l$	0.2501(14.7)	
			0.1074(12.3)
Total	1.6993	0.8730	

D연구

G연구 결과를 바탕으로 G연구와 동일한 국면의 조건
의 수로 수행한 D연구 결과는 다음 <Table 7>과 같다.
평가의 적절한 일반화가능도 수준은 연구자가 판단하며
(Kang & Lee, 2006), 대체로 0.8 이상의 높은 일반
화가능도 계수를 보이는 것을 확인하였다.

두 영역의 합성점수에 대한 일반화가능도 계수가 모
두 0.9 이상으로 높은 값을 나타냈다. 이에 비해
($p^* : a^*$) $\times r^*$ 설계에서의 일반화가능도 계수를 살
펴보면 볼과 후프 종목에서 비교적 낮은 값이 나타났고
(볼 0.7140, 후프 0.6912), 이는 볼과 후프 종목은 선
수의 지역 요인을 고려한 평가 상황의 신뢰도가 떨어진
다는 것을 의미한다. ($p^* : l^*$) $\times r^*$ 설계에서는 대부
분의 일반화가능도 계수가 0.8 이상의 값을 보였지만 리
본 종목의 계수가 0.7845로 낮게 나타났다. 이것은 리
본 종목에서 지도자의 명성 요인을 고려한 평가 상황의
신뢰도가 상대적으로 떨어진다는 것을 의미한다. 전반적
으로 선수의 지역이나 지도자의 명성 요인을 고려하지
않은 경우의 일반화가능도 계수가 가장 높은 수준을 보
였지만 이러한 요인을 고려한 경우에도 일부 종목을 제
외하면 어느 정도 높은 일반화가능도 계수를 산출하였다.

Table 7. Estimated generalizability coefficients for different designs

$p^* \times r^*$ design			
	Difficulty	Execution	Total
Ball	0.9600	0.9419	0.9758
Clubs	0.9756	0.9395	0.9806
Hoop	0.9656	0.9450	0.9779
Ribbon	0.9553	0.9624	0.9773
$(p^* : a^*) \times r^*$ design			
	Difficulty	Execution	Total
Ball	0.7303	0.6701	0.7140
Clubs	0.9664	0.9708	0.9718
Hoop	0.7028	0.6283	0.6912
Ribbon	0.7588	0.8649	0.8212
$(p^* : l^*) \times r^*$ design			
	Difficulty	Execution	Total
Ball	0.8723	0.8600	0.8715
Clubs	0.8726	0.8683	0.8796
Hoop	0.8795	0.8980	0.8967
Ribbon	0.7494	0.8119	0.7845

논 의

이 연구의 목적은 리듬체조 경기에서 나타나는 심판
평가의 신뢰도와 오차요인을 탐색하는 것이다. 이를 위
해 고등부 시니어 후프, 볼, 곤봉, 리본 4종목 평가 상황
에서 일반화가능도에 영향을 주는 요소들의 상대적 크기
는 어느 정도인가를 분석하였다.

연구문제와 관련하여 첫째, G연구 분석결과 모든 종
목에서 동일하게 난도영역과 실시영역에서 일반화가능
도에 영향을 주는 요소는 평가 대상인 선수로 나타났다.
이는 심판들이 선수들의 수행능력에 따라 타당하고 신뢰
로운 판정을 했다는 것을 의미한다. 그런데 D연구 결과
난도영역과 실시영역에 대한 일반화가능도 계수가 높게
나타나 신뢰도를 확보하였으나, 후프와 리본 등 몇 개의
이벤트에서는 일반화가능도계수가 낮게 나타나 신뢰성
에 문제가 있음을 보였다. 둘째, D연구결과 0.8이상의
일반화 가능도 계수가 나타나 일반적으로 신뢰로운 판정
을 보였으나 후프경기 판정은 선수의 지역에 따라 편향
된 점수를 주는 것으로 나타났다. 셋째, 리본경기 심판
판정은 지도자의 명성에 따라 편향된 점수를 주는 것으
로 나타났다.

이러한 연구결과와 관련하여 몇 가지 논의를 진행하
고자 한다. 첫째, 연구문제 1은 G연구결과 난도(D)와
실시영역(E)에서 선수의 분산성분이 가장 큰 값을 나타
냈다. 즉 이것은 심판이 선수의 경기력만을 보고 평가한
것으로 객관적이고 공정하게 채점을 했다는 것을 의미하
며 선행연구 Cho & Choi(2015)와 Lee et al.(2016)
의 연구결과와 일치한다.

이러한 결과는 연구에 참가한 심판들이 국내 3급 이
상의 심판자격을 소지한 숙련된 심판들이기 때문에 기인
한 것으로 판단된다. 심판들은 경기 전 심판 강습회를
통해 심판으로서의 기본 소양과 자질을 충분히 숙지한
다. 뿐만 아니라 국내와 국제경기에서는 각 영역 심판간
의 점수 차이가 심할 경우 심판장은 심판들의 판정을 조
율할 수도 있다. 때문에 심판들의 채점기준이 대부분 비
슷하여 큰 차이가 없게 된다.

반면, 선수와 심판의 상호작용의 분산성분 즉, 잔차는
난도영역에서 적게는 8.9%, 실시영역에서 많게는 20%
가 나타났다. 그리고 난도심판보다 실시심판의 상호작용

분산성분이 더욱 큰 것으로 나타났다. 이것은 선행연구 Cho & Choi(2015)의 연구결과와 일치한다. 이와 같이 난도영역보다 실시영역에서 심판과 선수의 상호작용 분산의 비율이 높다는 것은 실시영역에서 보다 더 심판의 판정이 선수의 경기력이 아닌 다른 어떠한 요인이 작용하고 있다는 것을 의미하기 때문에, 실시 영역에 영향을 주는 요소들과 관련된 후속연구가 필요하다.

D연구 결과, 선수의 지역, 지도자의 명성을 구분하였을 때 대부분 0.8 이상의 일반화가능도계수가 나타났다. 이것은 심판의 평가가 신뢰성이 있음을 의미한다. 그러나 선수의 지역과 지도자의 명성을 구분하였을 때 상대적으로 낮은 일반화계수가 산출되었다. 특히, 볼, 후프 실시영역은 다른 종목에 비해 낮은 수치였다. 이는 볼, 후프 종목 판정은 선수들의 소속지 요인을 고려할 경우, 그리고 리본종목은 지도자의 명성을 고려했을 경우 심판판정의 신뢰도가 상대적으로 낮았다는 것을 말한다. 이와 같이 동일한 자료라 할지라도 일반화가능도 분석에서 적용하는 분석 모형에 따라 일반화가능도 계수는 달라질 수 있다(Li & Brennan, 2007; Lee et al., 2015). 그러나 지역이나 지도자의 명성과 같이 실제 판정상황에 영향을 줄 수 있는 국면들을 무시하는 분석모형을 사용할 때 추정된 신뢰도는 편향된 추정치(biased estimates)일 수 있으므로(Li & Brennan, 2007), 심판판정의 일반성에 대한 해석에는 주의가 필요하다.

본 연구는 Cho & Choi(2015)의 연구 결과에서 잔차의 분산성분(최대 24.3%)이 상대적으로 선행연구들에 비해 높은 것에 주목하고, 이 연구에 적용된 분석모형이 판정상황에 대한 여러 국면들을 반영하지 못했을 것으로 추정하였다. 이들의 연구에서 제시된 일반화가능도 계수는 0.9755~0.9926의 범위를 나타내어 매우 높은 수준의 신뢰도를 보였다. 하지만 본 연구를 통해, 리듬체조 심판들은 선수의 지역이나 지도자의 영향을 받을 수 있으며, 이와 같은 국면을 연구에 포함할 경우 일반화가능도 계수는 달라질 수 있음을 유의해야 한다.

둘째, 연구문제 2와 관련하여 리듬체조 심판판정은 선수의 지역에 따라 편향된 점수를 주는 것으로 나타났다. G연구결과, 난도와 실시 영역은 유사한 결과를 보였으며 종목별로 상이한 결과를 나타냈다. 볼, 후프, 리본에서는 선수의 분산성분이 가장 높게 나타났으며 그 다

음으로 지역적인 요인이 높은 값을 나타냈다. 특히 곤봉 종목의 경우는 특정 지역에 대한 분산성분이 가장 높은 값을 나타냈다. 이러한 결과는 국제대회에서 심판들이 자신의 국적과 같은 자국 선수들에게 편향적으로 우호적인 평가를 했던 것처럼(Ansorge & Scheer, 1988; Ste-Marie, 1996) 국내경기에서 심판들은 같은 시도 선수들, 즉 지역적으로 가까워 보다 더 친밀감이 높을 수 있는 선수들에게 일부 종목에서는 높은 점수를 부여하는 경향이 있는 것으로 추론할 수 있다. 이러한 점은 Oh & Kim(2015)의 연구결과와 동일하다. 이러한 결과는 특정지역에 다소 밀집되어 있는 심판진 구성이 원인일 수 있다. 리듬체조 대회는 심판구성을 할 때 지역 제한의 규정이 없기 때문에 심판이 자신과 동일지역 출신 선수를 판정할 수 있다. 또한 선수와 지도자 심판들 대부분이 서울과 수도권에 몰려있기 때문에 서울과 수도권 출신 선수들이 우호적인 평가점수를 받을 수 있는 개연성이 있다. Kwak et al.(2016)이 주장했던 것과 같이 수도권을 제한 나머지 지역들은 고립되어 있어 새로운 정보 또는 경기관련 정보교환에 불리한 위치에 있다. 따라서 본 연구 결과처럼 일부 종목에서 심판들의 판정이 지역 요소에 영향을 받고 있는 것을 고려할 때, 심판배정이 수도권 중심의 집중된 구성일 경우에는 심판판정의 공정성이 훼손될 가능성이 있다.

셋째, 연구문제 3과 관련하여 리듬체조 심판판정은 지도자의 명성에 따라 편향된 점수를 주는 것으로 나타났다. 구체적으로 후프종목의 경우 지도자 명성의 분산성분이 난도(D) 44.5%, 실시(E) 46.1%로 가장 높은 값을 나타냈다. 즉, 심판판정이 선수의 경기력뿐만 아니라 지도자 명성에 의한 영향을 받고 있음을 알 수 있었다.

유명한 지도자의 선수들이 실제 수행능력보다 높은 점수를 받는 결과는 후광효과가 존재함을 말한다. 이러한 심판판정 오류는 Moon & Jung(2005)의 주장이 사례연구 이상임을 뒷받침 해준다. 즉 실력이 비슷한 선수의 경기력을 평가 할 때, 명성이 높은 지도자에게 배우는 선수는 조금 더 관대한 점수를 부여하고 명성이 낮은 지도자에게 배우는 선수의 점수는 엄격하게 평가한다는 것이다. 지도자들의 명성이 선수들의 경기결과와 관련이 있음을 알 수 있었다.

지도자의 명성 또는 후광이 심판판정에 영향을 미칠

수 있다는 결과는 지도자와 선수들이 비교적 소수인 점이 원인일 수 있다. 즉 리듬체조는 다른 종목에 비해 선수층과 지도자층이 얇고, 국내 경기의 경우 대부분의 지도자들이 심판 역할을 수행한다. 그렇기 때문에 심판들은 누가 대표지도자였는지, 잘 가르치는 지도자인지를 서로 잘 알고 있으며 또한 누가 영향력 있는 지도자인지를 체득한 상태에서 심판을 보게 된다. 따라서 선행연구들(Moon & Jung, 2005; Oh & Kim, 2015)이 지적한 것처럼, 심판판정은 지도자들의 명성 또는 사회적 관계에 영향을 받아 판정 오류를 범할 수 있다.

일반화가능도 이론을 적용하여 리듬체조 심판의 판정 결과를 분석한 Cho & Choi(2015)의 연구에서는, 설정한 분석모형에서 해석하지 못하는 전차의 최대분산성분 값은 24.3%였다. 본 연구는 선행연구에서 심판 판정에 영향을 주는 것으로 그 요인이 밝혀지지 않는 잔차의 영역에 선수의 지역요인과 지도자의 명성요인이 포함될 수 있음을 제시하였다. 일반화가능도 이론 분석에서는 왜 특정 종목에서 잔차의 분산 성분의 값이 높은 원인에 대해서는 구체적인 정보를 제공하지 않는다. 본 연구를 기반으로 심판 판정의 공정을 높이려는 측면에서 심판들의 판정이 지도자의 사회적 혹은 선수(부모)들의 사회적 관계와 관련이 있는지 실증적인 후속 연구가 필요한 시점이다.

결론

연구결과, 첫째, 난도와 실시영역에서 선수와 심판만을 고려한 분석모형 적용결과, 리듬체조 심판들의 판정은 선수의 분산분석 성분이 가장 컸다. 즉, 심판들은 선수의 실력에 따라 판정을 하고 있었다.

둘째, 난도와 실시영역에서 선수, 심판, 지역을 고려한 분석모형 적용결과, 리듬체조 심판들의 판정은 선수의 분산분석 성분이 가장 컸으나, 곤봉종목에서 지역의 분산 값이 가장 컸다. 이는 곤봉종목에서 선수의 지역에 따른 심판의 판정 편향성이 있을 수 있음을 의미한다.

셋째, 난도와 실시영역에서 선수, 심판, 명성을 고려한 분석모형 적용결과, 리듬체조 심판들의 판정은 선수의 분산분석 성분이 가장 컸으나, 후프종목에서 지도자

의 명성의 분산 값이 가장 컸다. 이는 후프종목에서 지도자의 명성에 따른 판정의 편향성이 있을 수 있음을 의미한다.

넷째, 일반화가능도의 신뢰도 분석에서 선수의 지역이나 지도자의 명성을 고려한 경우에는 모든 종목에서 비교적 낮은 신뢰도계수가 산출되었으나, 선수의 지역을 고려하지 않은 경우에는 일반화가능도 계수가 높은 값을 보였다. 이는 분석모형에 따라 신뢰도 해석에 주의가 필요함을 나타낸다.

요약하면, 이 연구는 리듬체조 심판들이 대부분 선수들의 수행에 근거하여 신뢰할 수 있는 판정을 하고 있음을 제시하고 있다. 그러나 일부 종목은 선수들의 출신 지역 또는 선수들의 지도자와 관련된 판정오류가 있다고 판단된다. 이러한 심판들의 판정오류를 개선하기 위해 몇 가지 실제적인 방법을 생각해 볼 수 있다. 첫째, 심판 판정은 고도의 전문성이 요구된다. 어떠한 경기든 심판 판정의 오차를 완전히 없애는 것은 현실적으로 불가능하다. 하지만 가능한 정확한 판정을 위해 즉, 의도적인 판정오류를 최소화하기 위해서는 시합 전·후 주기적인 역량강화 교육 또는 review meeting을 통해 전문성을 향상시켜야 하며 심판자격검정과정도 매우 엄격하게 진행해야 할 것이다. 둘째, 전문성 향상을 위해 국내와 국제 대회 심판 경험을 심화 시키도록 해야 한다. 심판역량강화 교육뿐만 아니라 많은 대회를 통해 심판 경험을 쌓도록 해야 한다.

현재 국내심판들은 심판 경험 기회가 매우 제한적이다. 국내 뿐 아니라 국제대회를 통한 심판 경험을 쌓을 수 있는 제도가 필요하다. 국내 심판들 중 국제대회에서 활동하고 있는 국제심판의 수는 매우 적다. 협회 추천을 통한 참가자격, 비용적인 문제 등 복잡한 절차로 진행되기 때문이다.

국제 강습회 및 워크샵은 다양한 국적의 심판들이 참가하여 경기에서 일어나는 문제점이나 다양한 정보를 나눌 수 있으며 시험을 통한 국제심판자격도 주어진다. 꼭 국제심판 자격의 목적이 아니더라도 국내와 국제심판 경험을 쌓을 수 있는 좋은 기회가 될 것이고 그러한 경험을 통해 판정 오류를 줄일 수 있을 것이다. 따라서 국제 강습회에 국내 심판들이 참가할 수 있도록 제도적인 방안 마련에도 힘써야 할 것이다. 이러한 노력은 선행연구

(Lee et., 2016)에서 심판관정의 문제점으로 지적될 수 있는 심판의 문화적 감수성 측면에서 국내 리듬체조 심판들의 판정능력을 국제적 수준으로 높이는 하나의 방안이 될 수 있을 것이다.

또한 지도자가 아닌 심판으로만 활동하는 전문적인 전임 심판제도가 마련된다면 판정오류를 줄일 수 있을 것이라 생각한다. 선수들의 출신지역이나 선수들의 지도자의 명성 등 경기력 외의 요인으로 인한 판정 오류를 없앨 수 있을 것이다. 전임심판은 의도적인 판정 오류에 대한 보다 막중한 책무성을 갖기 때문이다.

본 연구는 국내 리듬체조 대회의 심판 판정 자료를 기반으로 일반화가능도 분석을 활용하여 심판의 판정이 선수의 지역과 지도자의 명성에 영향 받을 수 있다는 것을 밝히 초기 연구로서의 의미가 있다. 현재 우리나라에서는 축구, 농구, 배구 등의 프로스포츠가 일 년 내내 운영되고, 심판의 오심은 언론에 크게 보도되기도 한다. 심판 판정의 질적인 개선을 위해 실증적인 연구가 필요한 시점(Thelwell et al., 2013)에 다양한 스포츠 영역에서 일반화가능도 분석을 활용한 심판 판정 관련 연구가 실천되어야 할 것이다.

참고문헌

- Ansorge, C. J. & Scheer, J. K. (1988). International bias detected in judging gymnastic competition at the 1984 olympic games. *Research Quarterly for Exercise and Sport*, 59(2), 103-107.
- Auweele Y. V., Boen F, De Geest A, & Feys J. (2004). Judging bias in synchronized swimming: Open feedback leads to nonperformance-based conformity. *Journal of Sport and Exercise Psychology*, 26, 561-571.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, 60, 556-560.
- Brennan, R. (2001a). *Generalizability theory*. New York: Springer-Verlag.
- Brennan, R. (2001b). *Manual for urGENOVA*. Iowa City, IA: Iowa Testing Programs, University of Iowa.
- Catarina, L., Lurdes, A. C., Elena, S. P., & Marta, B. A. (2016). The evaluation rules in the view of the rhythmic gymnastics judges. *Journal of Sports Science*, 4, 232-240.
- 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 Sport Science*, 17(3), 13-22.
- Findlay, L. C., & Ste-Marie, D. (2004). A reputation bias in figure skating. *Journal of Sport and Exercise Psychology*, 26, 154-166.
- Kang, A. N., & Lee, G. (2006). A generalizability theory approach to investigating the generalizability of performance assessment using student peer reviews. *Journal of Educational Evaluation*, 19(3), 107-121.
- Kim, K. S., Lee, G., & Kang, S. H. (2010). Analysis of error sources and estimation of reliability in a korean speaking Achievement. *Korean Language Education*, 21(4), 51-75.
- Kim, S. S. (2001). *Generalizability theory*. Seoul: kyoyookbook.
- Korea Institute for Curriculum and Evaluation. (2004). *Education assessment*. Seoul: Hakjisa.
- Kwak, J. H., Lee, H. J., & Lee, T. G. (2016). Professional community of coaching practice: Rhythmic gymnastics coaches' social network. *Korean Journal of Sport Science*, 27(4), 878-891.
- Lee, S. Y., Kim, S. Y., Kim, J. H., Baek, K. C., & Lee, B. Y. (2015). Analyses of the reliability of a preliminary creativity test using the multivariate generalizability theory. *Journal of Creativity Education*, 15(3), 83-107.
- Lee, T. G., Lee, H. J., & Yang, H. W. (2016). Analysis of judges' judging in figure skating. *Korean Journal of Sport Science*, 27(4), 756-769.
- Lee, T. G., & Yang, H. W. (2016). Analysis of error sources and estimation of reliability in peer review of forced connection method sports casting by applying generalizability theory. *Korean Journal of Sport Science*, 27(2), 345-361.
- Leskošek, B., Čuk, I., Pajek, J., Forbes, W., & Bučar-Pajek, M. (2012). Bias of judging in men's artistic gymnastics at the European championship 2011. *Biology of Sport*, 29 107-113.
- Li, D., & Brennan, R. (2007). *A Multi-group generalizability analysis of a large-scale reading comprehension test*. In annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Moon, W. J., & Jung, K. H. (2005). An ethnographic study on

- the decision determinants of taekwon-do referees. *Journal of Coaching Development*, 7(4), 59-72.
- New York Times (2013.07.10.). World body expels top london Olympics official in rhythmic gymnastics.
- Oh, J. S. & Kim, S. Y. (2015). A case study on judgments in women's apparatus gymnastics. *Journal of Coaching Development*. 17(3), 167-178.
- Plessner, H. (1999). Expectation biases in gymnastics judging. *Journal of Sport and Exercise Psychology*, 21, 131-144.
- Popovic, R. (2000). International bias detected in judging: Rhythmic gymnastics competition ay Sidney-2000 Olympic Games. *Physical Education and Sport*, 1(7), 1-13.
- Scheer, J. K., Ansoerge, C. J., & Howard, J. (1983). Judging bias by viewing contrived videotapes: A function of selected psychological variables. *Journal of Sport Psychology*, 5, 427-437.
- Sportschosun (2011.10.11.). '신수지 사태' 아름답지 못한 리듬 체조, 이대론 안된다.
- Ste-Marie, D. (1996). International bias in gymnastic judging: Conscious or unconscious influences? *Perceptual and Motor Skill*, 83(3), 963-975.
- Ste-Marie, D. M., & Valiquette, S.M. (1996). Enduring memory-influenced biases in gymnastic judging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1498-1502.
- Thelwell, R. C., Page, J. L., Lush, A., Greenlees, A., & Manley, A. J. (2013). Can reputation biases influence the outcome and process of making competence judgments of a coach? *Scandinavian Journal of Medicine and Science in Sports*, 23, e65-e73.

리듬체조 심판판정의 신뢰도 및 오차요인 분석

이태구 ·곽주현 ·양희원 ·이한주(연세대학교)

【목적】 이 연구의 목적은 일반화가능도 이론을 통해 리듬체조 심판판정의 신뢰도와 오차요인을 조사하는 것이다. 일반화가능도 이론은 연구자가 설정한 평가 상황에서 측정된 자료의 오차요인들을 포함한 측정모형을 근거로 오차요인을 원인별로 정량화하여, 판정상황에서 각 오차요인이 차지하는 상대적인 영향력을 파악(G연구)할 수 있으며, 이를 바탕으로 측정모형의 신뢰도까지 제시(D연구)할 수 있는 분석방법이다. **【방법】** 연구자료는 2016년 전국리듬체조 대회에 참가한 34명의 시니어부 선수들의 후프, 볼, 곤봉, 리본 종목의 판정 점수를 사용하였다. 각각 4명으로 구성된 난도(D)와 실시(E)심판의 채점 자료를 분석하였다. 선행연구(Kwak et al., 2016)를 기반으로 선수의 지역과 지도자의 명성을 포함한 분석모형을 설계한 후, 다변량 일반화가능도 이론을 활용하여 분석하였다. **【결과】** 연구 결과는 다음과 같다. G 연구결과, 첫째, 선수와 심판요인만을 고려한 측정모형에서, 난도와 실시영역 두 판정은 모두 선수의 분산성분이 가장 큰 값을 나타냈다. 둘째, 첫 번째 측정모형에 지역 요인을 포함한 분석에서는, 난도와 실시영역 두 판정은 모두 선수의 분산성분이 가장 큰 값을 나타냈으나, 곤봉 종목의 경우 지역의 분산성분이 가장 높은 값을 보였다. 셋째, 첫 번째 측정모형에 지도자의 명성 요인을 포함한 분석에서는, 난도와 실시영역 두 판정은 모두 선수의 분산성분이 가장 큰 값을 나타냈으나, 후프 종목의 경우 지도자의 명성의 분산성분이 가장 높은 값을 보였다. D 연구 결과, 넷째, 선수의 지역이나 지도자의 명성을 고려하지 않은 경우의 일반화가능도 계수가 가장 높은 값을 보였으나, 이러한 요인들을 반영하였을 때는 일부 종목들에서 비교적 낮은 신뢰도계수가 산출되었다. **【결론】** 본 연구는 실제 심판들이 판정한 데이터를 기반으로 일반화가능도 분석을 통해 심판의 편향적 판정을 초래하는 요소들에 대한 영향력을 분석한 연구이다. 연구결과를 바탕으로 리듬체조 심판들의 지역 및 지도자의 명성 요인에 영향을 받는 심판판정의 문제와 원인, 그리고 개선을 위한 대안을 논의하였다.

주요어: 일반화가능도 이론, 리듬 체조, 심판판정, 신뢰도, 오차요인, 편향적 판정