


Analysis of Judges' Judging in Figure Skating

Tae-Koo Lee, Han-Joo Lee, Jae-Eun Chung, & Hee-won Yang*

Yonsei University

The purpose of this study was (1) to analyze judges' evaluation on figure skating performance and statistical analysis and (2) to make recommendation to improve judges' performance. Data were 62 figure skaters' scores from Senior Woman 1 Group Part at 2015 The National Figure Skating Championship in Korea. Data of presentation part in Short Program were analyzed. Presentation part consists of skating skills, transitions/linking footwork/movement, performance/execution, choreography/composition and interpretation. Nine judges evaluate skater's presentation with the score 10.0 for each factor. Generalizability theory, descriptive statistics and ANOVA were utilized. Results showed that generalizability coefficient of presentation part was over 0.9, therefore stable reliability was secured. The error source about players has more significant impact to evaluation than other error sources. Generalizability coefficient was stable when reducing the number of judges up to 2. Also, judges generally represented aspects of similar evaluation, but different aspects appeared on a few judges' assessment in the components of presentation. Discussions were provided in terms of the reliability of the judging method for the presentation part of figure skating and the effective measurement condition.

Key words: Generalizability Theory, Figure Skating, Presentation Part, Reliability, Error Source, Tendency of assesment 

서론

피겨 스케이팅, 체조와 같은 경기들에서 심판들은 선수들의 연기를 관찰하고 채점을 한다. 선수들의 연기에 대해 심판들은 전문성을 가지고 공정한 채점을 하지만, 심판의 판정과 채점은 종종 논쟁을 만들었다. 예를 들어, 2014년 소치 동계올림픽 피겨 스케이팅에서 김연아

선수가 개최국 러시아 선수인 소트니코바에 밀려 은메달을 획득한 것이 대표적이다. 실제로 여러 스포츠 종목에서 심판 판정과 관련한 일부 선행연구들(Choi & Choi, 1999; Han & Kim, 2009; Kim & Kim, 2000)에서는 선수들이 심판판정이 공정하지 못하다는 인식을 가지고 있음을 보여주었다. 공정한 판정을 위해 여러 종목은 비디오 판독의 도입이 대안으로 제시되고 있으며, 야구, 농구 등의 일부 종목들에서 이미 도입하여 적용되고 있다(Cha et al., 2015; Park & Park, 2012; Seo, 2011). 피겨 스케이팅 종목에서는 2002년 솔트레이크 올림픽 페어 종목에서 심판의 부정으로 공동 금메달이 수여되면서, 그 이후로 비디오 판독이 도입되었다.

스포츠경기에서 심판판정은 선수들의 경기력에 영향을 주고, 궁극적으로는 승패나 대회에서의 메달 색깔을 결정할 수도 있기 때문에 중요한 연구주제이다. 그렇기

논문 투고일 : 2016. 07. 28.

논문 수정일 : 2016. 08. 23.

계재 확정일 : 2016. 09. 08.

* 저자 연락처 : 양희원(yhw1627@naver.com).

* 본 논문은 '2016 The 28th International Sport Science Congress in Commemoration of The 1988 Seoul Olympic Games'에서 발표된 'Do judges grade Kim Yuna's presentation in figure skating fairly? Investigation of error sources, reliability and tendency of assessment using generalizability theory and statistical analysis'를 수정·보완한 논문임.

때문에 연구자들은 스포츠 심판들을 연구주제로 하여 다양한 종목에서 연구를 실행해 왔다. 다양한 선행 연구들은 경기 상대가 있는 축구, 배구, 태권도 등에서 심판의 판정실책인 오심의 원인 및 경향을 분석하거나(An, 2014; Kim & Jeon, 2014; Park, 2003) 체조/리듬 체조, 에어로빅, 피겨스케이팅과 같이 심판의 채점 결과를 분석하는 연구(Ahn, 2014; Ansoerge & Scheer, 1986, 1988; Cho & Choi, 2015; Findlay & Ste-Marie, 2004; Kim et al., 2011)로 구분할 수 있다.

지금까지 심판이 선수의 일련의 연기를 관찰한 후 채점한 평가결과를 분석한 선행연구들은 주로 기술통계나 추리통계를 통해 심판들의 판정오류를 확인하였다. 예를 들어, Ansoerge & Scheer(1988)는 *t*-test를 통해 1984년 LA올림픽 체조경기에서 심판의 국적과 동일한 선수를 채점한 평가점수와 동일한 선수에 대하여 국적이 다른 나라 심판이 채점한 평가점수를 비교한 결과, 심판들은 자신의 국적과 동일한 선수들의 연기에 대하여 다른 국적 선수들에 비해 통계적으로 유의하게 높은 점수를 부여하고 있음을 확인하였다. 또한 Findlay & Ste-Marie(2004)는 캐나다 피겨 스케이팅 연맹 소속의 12명의 전문심판들을 대상으로 14명의 여자 피겨선수들의 연기를 편집한 동영상을 보고 채점하게 하였다. 연구결과, 심판들은 피겨의 쇼트 프로그램의 Technical Merit에서 명성이 있는 선수들에게 높은 점수를 주고 있었다. 이와 같이 기술통계나 추리통계의 방법을 통한 심판판정 결과에 대한 분석 연구들은 심판들의 판정오류를 밝히는데 효과적이었다. 그렇지만, 이러한 연구들은 발견된 문제들을 해결하거나 예방할 수 있는 측정학적인 조건과 방법들을 구체적으로 제시하지 못하는 한계가 있었다.

이러한 제한점을 극복하기 위해 최근에는 관찰자가 판정한 결과에 대하여 오차변인, 즉 관찰자 오차를 통계적으로 분석하여 관찰자에게 가장 크게 영향을 준 요인들을 분석하고 이를 기초로 신뢰성이 높고 효율적인 측정조건을 제시할 수 있는 통계기법인 일반화가능도 이론(Generalizability theory)이 도입되었다. 일반화가능도 이론을 이용하면 관찰자인 심판의 판정 오차를 분석의 초점에 두면서, 심판들의 판정에 대한 신뢰도와 오차를 요소별로 영향력을 밝혀낼 수 있을 뿐만 아니라, 신뢰롭고 효율적인 측정 조건도 제시할 수 있는 장점이 있

다. 예를 들어, Cho & Choi(2015)는 국내의 리듬체조 대회에서 26명의 선수들을 대상으로 8명의 심판이 후프와 리본종목 선수 연기 즉, 난도와 실시영역을 채점한 평가점수를 일반화가능도 이론을 통해 분석하였다. 연구결과, 심판들은 후프와 리본종목의 난도와 리본종목의 실시 영역에서 신뢰롭게 채점을 하고 있었으나, 후프는 신뢰롭게 채점을 하지 못하고 있어 이 부분에서 심판의 재교육이 필요함을 제시하였다. 그리고 측정 조건과 관련하여 숙련된 심판들은 후프의 난도와 실시 영역에서는 2명 이상, 그리고 리본의 난도와 실시 영역에서는 1명 이상이 되어도 신뢰로운 측정이 가능함을 제시하였다. 이렇게 일반화가능도 이론을 통한 심판 채점 결과의 분석 연구는 심판판정 논란을 개선할 수 있다는 점에서 의미가 있다. Cho & Choi(2015)의 연구처럼 일반화가능도 이론은 심판들의 오차요인의 상대적 영향력과 효율적인 측정조건을 제시하여, 심판 판정을 개선하기 위한 대안을 제시할 수 있기 때문이다.

피겨 스케이팅 종목에서 심판판정관련 연구들은 외국의 연구자들(Findlay & Ste-Marie, 2004; Zitzewitz, 2014)뿐만 아니라 국내에서도 1990년대 후반 이후로 피겨 종목에 대한 관심이 높아지면서 연구자들의 주요한 연구 주제가 되어 왔다(Lee et al., 2014). 관련 연구들은 심판판정 결과에 통계적 분석을 통해 판정 경향 및 판정의 객관성을 분석하거나(Kim, 2003; Cho & Chun, 1997; Shin et al., 2010), 심판판정제도의 변경에 따라 영향을 받을 수밖에 없는 코칭방법/전략(Anh, 2003)과 유발되는 선수들의 스트레스관련 연구(Shin & Yoon, 2013)도 진행되었다. 그러나 이러한 선행연구는 주로 고전 검사이론 및 문항반응이론을 적용한 연구방법이 적용되었으며, 심판의 판정결과를 기초로 일반화가능도 이론을 적용하여 심판들의 오차요인의 상대적 영향력과 효율적인 측정조건을 제시하는 연구는 시도되지 않고 있다.

국제스케이팅연맹(ISU) 총회는 피겨 스케이팅에 실명 채점제 도입을 결정하였다. 실명 채점제는 심판들 각자의 채점점수가 공개된다는 점에서 심판들의 책무 수행에 대한 정보를 제공한다. 다시 표현하면, 심판들이 적절한 수행을 하고 있는지 심판 개개인의 채점 특성 및 경향에 대한 정보를 제시한다. 따라서 심판 개개인의 채점 경향에 대한 정보 분석은 피겨 스케이팅 종목 심판판정

개선을 위해 유용한 정보를 제시할 수 있다. 그러나 Cho & Choi(2015)의 연구처럼 일반화가능도 이론만을 적용한 선행연구들은 심판 개인들의 채점에 관한 정보는 제시하지 못하는 제한점을 보이고 있다. 따라서 이 연구의 목적은 피겨 스케이팅 연기영역 심판 평가의 신뢰도와 오차요인을 탐색하고, 채점 경향을 바탕으로 심판 평가 개선을 위한 대안을 제시하는 것이다. 이를 위한 구체적인 연구문제는 다음과 같다.

- 첫째, 피겨 스케이팅 연기영역 평가에 영향을 주는 평가상황 요소들의 상대적인 영향력 크기는 어느 정도인가?
- 둘째, 피겨 스케이팅 연기영역 평가에서 적정수준의 일반화가능도를 확보할 수 있는 효율적인 측정 조건은 어떠한가?
- 셋째, 피겨 스케이팅 연기영역 평가에서 심판 개인 및 연기 평가 세부영역 별 평가 양상은 어떻게 다른가?

연구방법

연구 자료

이 연구의 연구대상은 2015 전국남녀 피겨 스케이팅 랭킹대회(12/4-12/6, 서울 목동 빙상장)의 시니어 여자 싱글 1그룹 규정종목인 쇼트 프로그램에 참가한 62

명으로 이들의 채점결과를 대한빙상연맹 홈페이지에서 다운받아 원자료로 수집하였다. 피겨 스케이팅 종목에서 심판진은 선수들의 연기의 필요 조건들을 판정하는 3명의 Technical Panel, 연기의 기술과 구성을 판정하는 9명의 Judge와 이를 총괄하는 1명의 Referee로 구성되며, 본 연구에 참여한 심판들은 모두 대한빙상경기연맹 소속 1급 이상의 자격을 갖춘 심판들이었다.

쇼트프로그램은 2분 40초의 연기 시간에 이루어진 선수의 수행을 채점하는 것으로, 채점 영역은 수행(Executed Elements)점수와 프로그램 구성(Program Components) 점수로 구성되고 9명의 심판들이 채점한다. 수행점수는 피겨 스케이팅 수행요소에서 기술적인 영역과 관련된 것으로 ISU에서 규정한 7가지 기술 수행요소들(세 종류의 점프와 스핀 그리고 한 종류의 스텝)에 대해 연기의 정확성과 난이도를 평가하는 것이다. 구성점수는 선수가 수행하는 모든 표현의 수준, 즉 예술적인 측면에서 선수의 연기(presentation)를 평가한다. 구성요소는 총 다섯 가지 요소로 구성되는데, 스케이팅 스킬(Skating Skills), 트랜지션/연결 풋워크 및 동작(Transitions/Linking Footwork/Execution)과 연기/수행(Performance/Execution)은 피겨 스케이팅의 예술성과 기술성이 복합된 성격을 가지고 있으며, 안무/구성(Choreography/Composition)과 음악적 해석/타이밍(Interpretation)은 전형적인 예술적 성격을 가지고 있다. ISU에서 제시하는 구성요소들의 구체적인 의미는 <Table 1>에 제시하였다.

Table 1. Definition of the Program Components

Factor	Definition
SS(C1)	Overall skating quality, edge control and flow over the ice surface demonstrated by a command of the skating vocabulary(edges, steps, turns etc.), the clarity of technique the use of effortless power to accelerate and very speed. Varied use of power/energy, speed and acceleration.
TR(C2)	The varied and/or intricate footwork, positions, movements and holds that link all elements, including the entrances and exits of those elements.
PE(C3)	Performance: involvement of the Skater/Pair/Couple physically, emotionally and intellectually as they translate the intent of the music and choreography. Execution: quality of movement and precision in delivery. This includes harmony of movement in Pair Skating and Ice Dance.
CH(C4)	An intentional, developed and/or original arrangement of all types of movements according to the principles of proportion, unity, space, pattern, structure and phrasing.
IN(C5)	The personal and creative translation of the rhythm, character and content of music to movement on ice.

SS=Skating Skills, TR=Transitions/Linking Footwork and Movement, PE=Performance/Execution, CH=Choreography/Composition
IN=Interpretation of the Music/Timing

심판(judge)들은 수행점수와 구성점수를 각 요소 영역별로 채점을 하는데, 구성점수는 선수 연기에 대한 심판의 예술적 평가를 의미하므로, 구성점수는 심판에 의한 주관적 운동기능평가의 성격을 가지고 있다. 따라서 일반화가능도 이론에서 주목하는 관찰자 오차가 존재할 것으로 예견되기에 구성점수를 연구 자료로 이용하였다.

채점 방법

ISU(2014) 채점 규정에 의해 심판들은 선수가 연기를 마친 후 0.25씩 증가하는 0.25~10까지의 등급으로 평가하는데, 이는 절대평가의 특성을 지닌다. 한 선수는 다섯 가지 구성요소에 따라 9명의 심판들이 채점한 구성점수를 얻게 된다. ISU(2014)에서 제시하는 구성요소의 채점단계는 <Table 2>와 같으며, 평가 등급은 동간·비율척도의 성격을 갖는다. 각각의 단계에서 0.25씩 증가하는 증가점수는 심판들이 한 단계와 그 다음 단계의 수준이 섞여있는 연기에 대해 판정할 때 적용한다.

Table 2. Degrees of the Program Components

Point	Meaning
less than 1	extremely poor
1	very poor
2	poor
3	weak
4	fair
5	average
6	above average
7	good
8	very good
9~10	outstanding

자료 수집 및 분석 방법

자료 수집은 대한빙상경기연맹 홈페이지를 활용하였다. 대한빙상경기연맹에서 주관하는 국내의 모든 피겨대회 결과 및 심판 채점결과는 대회 종료 후, 대한빙상경기연맹 홈페이지에 그 결과를 게시하고 있다. 따라서 연구자들은 홈페이지에서 해당 대회의 참가선수별 심판 채

점표를 수집하여 분석에 활용하였다.

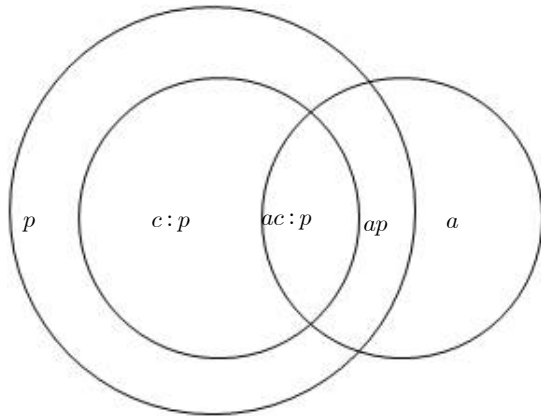
자료 분석에 사용한 방법은 일반화가능도 이론에 의한 분석과 기술통계 분석 및 일원배치 분산분석(one-way ANOVA) 등이다. 일반화가능도 이론은 측정상황에서 발생할 수 있는 여러 오차 요인에 대한 설명이 불가능한 고전검사이론의 단점을 극복하기 위한 개념적 틀을 제공한다. 즉 다중오차요인(multiple sources of error)을 동시에 분석하고, 측정된 점수에 대한 오차요인의 상대적인 영향력을 산출하여, 일반화가능도 계수와 함께 측정의 상황을 결정하는 연구자에게 안정적인 점수를 얻기 위한 측정조건(수준)을 제시함으로써 신뢰도 추정과정을 한 단계 향상시킨 것이다(Brennan, 2001).

일반화가능도 이론은 G연구(일반화연구)와 D연구(결정연구)로 구분된다. G연구에서는 분산분석 결과 얻어지는 각 분산성분들의 제곱평균(mean square)으로 관찰점수 분산을 구성하는 독립적인 분산성분 추정값을 산출하여 오차요인의 상대적 영향력을 제시한다. D연구에서는 G연구 결과를 바탕으로 일반화하고자 하는 전집(universe)을 규정하여 일반화가능도 계수(generalizability coefficient)를 산출하고, 각 측정국면의 조건에 따라 일반화가능도 계수 산출을 반복하여 적정 수준의 신뢰도를 확보하기 위한 효율적인 측정구조를 탐색한다.

이 연구는 연구대상인 피겨 스케이팅 선수들을 심판들이 세부요소에 따라 평가하였으므로, 이를 일반화가능도 이론 분석 체계로 표현하면 $a \times (c:p)$ 설계에 해당한다. 이 설계는 심판(p)들이 다섯 개의 세부요소(c)에 따라 평가하며, 모든 선수(a)들이 이에 따라 평가받는 구조이다. 이 중 선수(a)를 측정대상(object of measurement)으로, 심판(p)을 임의효과 국면(random facet)으로 설정하였다. 세부요소는 5개로 정해진 국면이므로, 고정효과 국면(fixed facet)으로 설정하였다(Fig. 1).

이와 같은 설계에 의하여 본 연구에서는 심판판정에 영향을 주는 요소로서 선수(a), 세부요소($c:p$), 심판(p), 선수와 심판의 상호작용(ap), 잔차($ac:p$)의 총 여섯 가지 분산성분(variance component)이 고려된다.

D연구는 G연구와 동일한 설계인 $a \times (C:P)$ 의 설계를 적용하였고, 측정국면 중 임의효과 국면인 심판의 표집 수를 조절하여 실제 평가상황에서 적정 수준의 일반화가능도를 유지하기 위한 효율적인 측정구조는 어떠



a: athlete effect
c: component effect
p: judge effect

Fig. 1. Venn diagram for $a \times (c:p)$ design

한지 탐색하였다.

D연구는 G연구와 동일한 설계인 $a \times (C:P)$ 의 설계를 적용하였고, 측정국면 중 임의효과 국면인 심판의 표집 수를 조절하여 실제 평가상황에서 적정 수준의 일반화가능도를 유지하기 위한 효율적인 측정구조는 어떠한지 탐색하였다.

심판의 판정이 심판 개인별로, 또는 심판이 판정하는 세부요소별로 어떻게 다른지 파악하기 위하여 각 심판별, 세부요소별 평균과 표준편차를 산출하여 비교하였다. 또한 각 심판별, 세부요소별 판정이 통계적으로 유의미한 차이를 나타내는지 확인하기 위하여 일원배치 분산분석을 실시하였다. 일원배치 분산분석 후 구체적으로 어떤 심판들 간, 또는 어떤 세부요소 간의 차이에 기인하여 평균 차이가 발생하였는지 살펴보기 위하여 사후비교분석을 실시하였다. 이 때 집단 간 사례수(선수 수)가 동일하므로, 사후비교분석 방법 중 Tukey 방법을 사용하였다.

일반화가능도 설계의 분산성분을 추정하는 G연구 분석과, 이를 바탕으로 오차 국면의 수를 조정함으로써 효율적인 측정 구조를 탐색하기 위한 D연구 분석을 위해서 GENOVA(Crick & Brennan, 1983) 컴퓨터 프로그램을 사용하였다(부록 참고). 각 심판별, 세부요소별 기술통계 분석 및 일원배치 분산분석에는 SPSS 프로그램을 사용하였다.

연구 결과

일반화가능도 이론 G연구

일반화가능도 이론에 의한 G연구 결과 전집점수(universe score) 분산과 오차점수 분산 성분의 크기를 상대적으로 비교하여 각 요인이 관찰점수(observed score) 분산에 미치는 영향력을 나타내는 분산성분 추정치는 <Table 3>과 같다.

측정대상인 선수의 분산성분($\hat{\sigma}^2(a)$)이 가장 큰 것으로 나타났고(0.6133), 다음으로 큰 분산성분은 선수와 심판의 상호작용($\hat{\sigma}^2(ap)$)이었다(0.0954). 심판($\hat{\sigma}^2(p)$) 및 세부요소($\hat{\sigma}^2(c:p)$)의 분산성분은 비교적 작았으나(0.0096, 0.0275), 세부요소의 분산성분이 심판의 분산성분보다 약 3배 큰 것으로 나타났다. 이 결과는 선수의 실력 외에 선수 각각의 점수에 가장 큰 영향을 미치는 요소가 선수와 심판의 상호작용이라는 것이다. 즉 선수에 따라 심판의 판정이 다른 양상을 보이며 선수의 점수에 영향을 주었다고 할 수 있다. 그 밖에 측정대상인 선수의 분산성분과 잔차($\hat{\sigma}^2(ac:p)$)를 제외한 분산성분 중 큰 비율을 차지하는 심판이 채점하는 세부요소에 따른 차이 역시 선수의 점수에 영향을 주고 있었다.

Table 3. ANOVA for $a \times (c:p)$ design

Effect (α)	df(α)	SS(α)	MS(α)	$\hat{\sigma}^2(\alpha)$	percentage
a	61	1712.6379	28.0760	0.6133	79.4
p	8	27.5575	3.4447	0.0096	1.2
c:p	36	62.3355	1.7315	0.0275	3.6
ap	488	232.7426	0.4769	0.0954	12.3
ac:p	2196	59.0116	0.0269	0.0269	3.5
Total	2789			0.7726	100.0

df(α): Degree of freedom for facet α

SS(α): Sum of squares for facet α

MS(α): Mean square for facet α

$\hat{\sigma}^2(\alpha)$: Estimated G-study variance component for facet α

일반화가능도 이론 D연구

G연구 결과 산출된 분산성분 추정치를 바탕으로, 고정효과 국면인 세부요소 수를 5개로 고정시키고 심판의 수를 변화시켜 일반화가능도에 미치는 영향을 알아보는 D연구를 수행하였다. 피겨 스케이팅의 쇼트 프로그램 체점의 경우 9명의 심판이 참여하는 것이 일반적이거나, 국내 대회에서는 여건에 따라 더 적은 수의 심판을 두고 체점을 하는 경우도 있다. 9명보다 많은 수의 심판이 참여하는 것은 대회의 효율성을 떨어뜨릴 수 있음을 고려하여 심판 수를 9명 이하에서 변화시켰다. 각 조건에서의 D연구 표집 수 및 분석 결과는 <Table 4>에 제시하였다.

Table 4. D-study for $a \times (c:p)$ design in different measurement conditions

no. of panels	$\hat{\sigma}^2(\tau)$	$\hat{\sigma}^2(\delta)$	$E\hat{\rho}^2$
1	0.6133	0.0954	0.8654
2	0.6133	0.0477	0.9279
3	0.6133	0.0318	0.9507
4	0.6133	0.0239	0.9626
5	0.6133	0.0191	0.9698
6	0.6133	0.0159	0.9747
7	0.6133	0.0136	0.9783
8	0.6133	0.0119	0.9809
9	0.6133	0.0106	0.9830

$\hat{\sigma}^2(\tau)$: Estimated universe score variance

$\hat{\sigma}^2(\delta)$: Estimated relative error variance

$E\hat{\rho}^2$: Estimated generalizability coefficient

이 연구의 자료수집에 사용한 조건인, 세부요소 5개 및 심판 9명일 때의 일반화가능도 계수는 0.9830으로 매우 높게 나타났다. 또한 심판 수가 9명 이하로 감소할 수록 일반화가능도 계수가 낮아지는 것을 확인하였다 <Fig. 2>.

평가의 적절한 일반화가능도의 수준은 연구자가 판단하며(Kang & Lee, 2006), 이 연구의 자료수집에 사용된 조건에서의 일반화가능도 계수가 0.9 이상으로 나타났다으므로 0.9 이상의 일반화가능도 계수를 유지하기 위한 상황을 확인하였다. 그 결과 5개의 세부요소를 유

지하는 경우 심판이 최소 2명 이상이 필요함을 확인할 수 있다. 그러나 심판이 1명인 경우의 일반화가능도 계수 역시 0.8654로 높은 수준이었다.



Fig. 2. Plot of Generalizability Coefficient Using Univariate Generalizability Theory

심판 및 세부요소별 기술통계량과 평균차이

심판 평가의 개선을 위하여 각 심판 및 세부요소별 선수의 점수 평균과 표준편차를 산출하여 이를 비교하였다. 또한 일원배치 분산분석을 사용하여 각 심판 및 세부요소별 평균 간 차이에 대한 통계적 유의성을 검증하였다. 심판 및 세부요소별 평균과 표준편차, 일원배치 분산분석 결과는 <Table 5>와 같다. 일원배치 분산분석 후 구체적으로 어떤 심판들 간, 또는 어떤 세부요소 간의 차이에 기인하여 평균 차이가 발생하였는지 살펴보기 위하여 사후비교분석을 실시하였다. 이 때 집단 간 사례 수(선수 수)가 동일하므로 Tukey 방법을 사용하였다.

가장 높은 평균 점수를 준 심판은 P3이고, 가장 낮은 평균 점수를 준 심판은 P9이다. 점수의 표준편차가 가장 크게 나타난 심판은 P5이고, 표준편차가 가장 작게 나타난 심판은 P4이다. F검정 결과 심판 및 세부요소별 평균 간 차이는 통계적으로 유의하였다(심판 평균 $F=4.635$, $p<0.001$ / 세부요소 평균 $F=18.535$, $p<0.001$). Tukey 분석결과, 심판 P9의 판정이 심판 P2, P3, P4, P6의 판정과 통계적으로 다른 차이를 보이는 것, 그리고 세부요소 C1과 C2, C3가 통계적으로 다른 차이를 보이는 것이 F검정의 통계적 결과에 영향을 주는 것으로 나타났다.

Table 5. Descriptive statistics and ANOVA results

	Mean	SD	F	Tukey	
P1	4.41	0.808			
P2	4.56^b	0.790			
P3	4.59^b	0.725			
P4	4.54^b	0.619			
Judge	P5	4.43	2.523	4.635***	a<b
	P6	4.49^b	1.031		
	P7	4.45	0.985		
	P8	4.39	0.887		
	P9	4.24^a	0.794		
Component	C1	4.59^c	0.800	18.535***	a<b<c
	C2	4.20^a	0.832		
	C3	4.43^b	0.851		
	C4	4.51	1.319		
	C5	4.55	0.883		

*** p<0.001

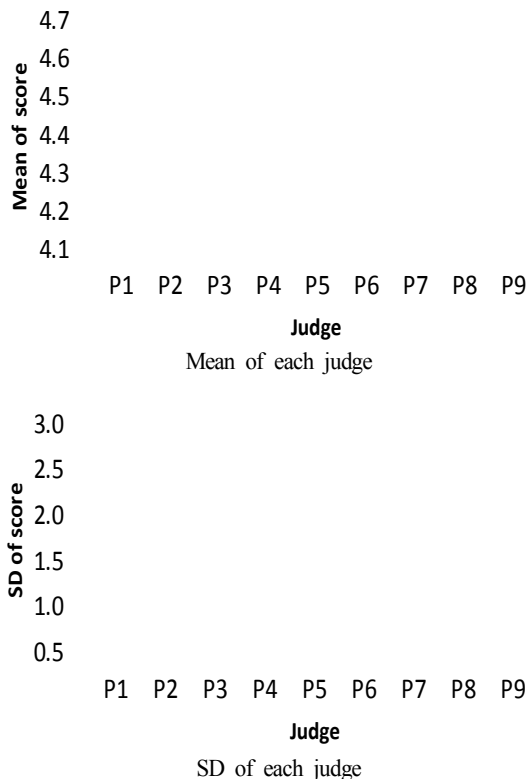


Fig. 3. Plot of Mean and Standard Deviation of each Judge

심판별 평가 점수에 대한 기술통계량을 살펴보았을 때 가장 높은 평균 점수를 준 P3 심판은 상대적으로 관대하게 판정한 것으로 해석할 수 있다(Fig. 3). 또한 P5 심판은 다른 심판들에 비해 매우 다른 표준편차를 보였다. 즉 P5 심판이 판정한 점수는 다른 심판들에 비해 넓게 산포되었다. 세부요소별 기술통계량에서 C2 세부요소는 다른 세부요소에 비해 0.23점 이상 낮은 평균 점수를 보였으며, C4 세부요소의 표준편차가 가장 크게 나타났다(Fig. 4.).

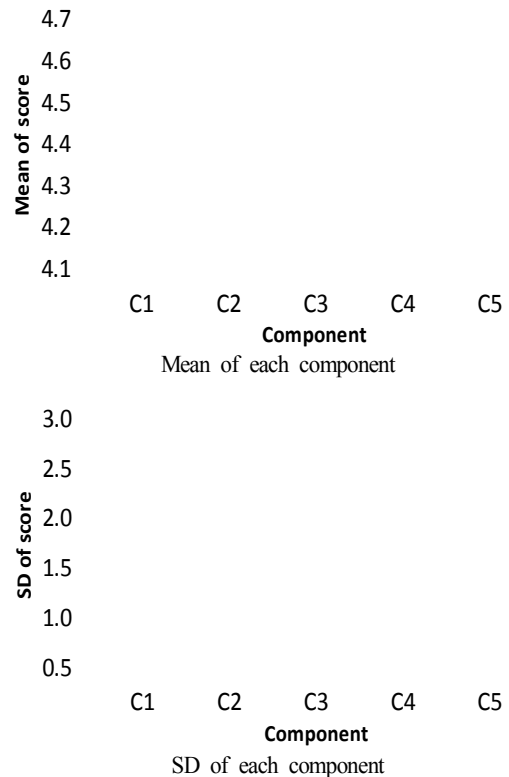


Fig. 4. Plot of Mean and Standard Deviation of each Component

논 의

이 연구의 목적은 일반화가능도 이론과 기술통계 분석을 사용하여 피겨 스케이팅 연기영역 평가의 신뢰도와 오차요인을 탐색하고, 심판 평가 개선을 위한 대안을 제

시하는 것이다. 이를 위해 시니어 여자 싱글 1그룹 규정 종목인 쇼트 프로그램에 참가한 선수들의 연기를 9명의 심판들이 채점한 결과를 분석하였다. 연구결과 첫째, G 연구 결과, 구성요소 채점 결과에 기여하는 오차 요인들의 상대적 영향력은 선수의 분산성분이 가장 컸지만, 선수와 심판의 상호작용의 분산성분은 전체 영역에서 12.3%를 차지하였다. 둘째, D연구 결과, 구성요소 채점을 위해 0.9 이상의 일반화가능도 수준을 유지하기 위해서는 최소 2명 이상의 심판이 필요하였다. 셋째, 각 심판 및 구성요소 세부요소들에 대한 기술통계량과 평균 차이 분석 결과, 일부심판들은 상대적으로 관대하거나 엄격하게 채점을 하였으며, 구성요소들 중 심판들이 채점한 C2의 평균점수는 가장 낮았으며, C4의 표준편차는 상대적으로 가장 크게 나타났다. 이러한 결과를 바탕으로 연구 문제와 관련하여 논의를 하면 다음과 같다.

첫째, 피겨 스케이팅 연기영역 평가에서 일반화가능도에 영향을 주는 평가 상황 요소들의 상대적인 영향력 중 가장 큰 요소는 선수요소였다. 이는 심판들은 선수들의 수행 능력에 따라 평가를 했다는 것을 의미하기 때문에, 심판들은 선수의 연기 수준에 따라 평가를 한 것으로 추론이 가능하다. 본 연구에 참여한 피겨 심판들은 1급 이상의 자격을 갖춘 심판들이기 때문에, 심판들은 기본적으로 선수들의 실력에 따라 채점을 하고 있었음을 본 연구결과는 보여주고 있다. 이러한 연구 결과는 문항 반응이론 분석을 통해 피겨 스케이팅 심판들의 채점 결과를 분석한 Kim(2003)의 연구에서 심판들의 채점의 엄격성이 전체적으로 일관되었다는 결과와 유사한 결과이며, 일반화가능도 이론을 적용하여 리듬체조 종목의 평가점수를 분석한 선행연구(Cho & Choi, 2015)의 결과와도 일치하는 것이다.

하지만 선수와 심판의 상호작용도 전체 분산성분의 12.3%를 차지하여 그 영향력이 선수 요소를 제외한 다른 요소들에 비해 상대적으로 컸다. 관찰자가 다른 대상을 평가하는 상황에서 일반화가능도 이론을 적용하여 분석할 때, 평가 대상이 선수이거나 학생일 경우, 평가 상황 요소들의 상대적인 영향력들 중 가장 큰 요소는 선수이거나 학생이어야 한다. 이러한 경우 평가자의 판정이 공정한 것으로 추론할 수 있다. 하지만, 그 밖의 요소들과 관련한 영향력의 분산성분의 비율이 상대적으로 높을

때는 심판의 평가가 대상의 실력만이 아닌 다른 요소의 영향을 받고 있음을 나타낸다. 일반화가능도 이론을 활용하여 리듬체조 종목의 평가점수를 분석한 Cho & Choi(2015)의 연구에서 선수요소의 분산분석 추정치들은 후프 종목 실시(E)에 대한 G연구를 제외하면, 각각 88.2%, 94.0%, 93.2%였으며, 선수와 심판의 상호작용의 분산성분은 각각 11.1%, 5.6%, 6.8%이었다. 이러한 결과는 본 연구에 참여한 피겨 스케이팅 심판들은 Cho & Choi(2015)의 연구에 참여한 리듬체조 심판들보다 선수와 심판의 상호작용의 영향이 컸음을 보여주는 결과이다. 이 결과는 주관적 판정에 의해 선수들의 순위가 결정되는 피겨 스케이팅 종목에서 심판의 판정이 선수의 실력 외적인 부분에 따라 달라질 수 있음을 의미한다. 선행연구들에서 판정자의 관찰자 오차가 발생할 수 있는 피겨 스케이팅이나 체조의 심판들은 선수들의 실력 요인 외에 국적(Ansorge & Scheer, 1986), 문화적 배경(Zitzewitz, 2014), 명성(Findlay & Ste-Marie, 2004), 연기순서(Ansorge et al., 1978), 또는 무의식적 선호(Ste-Marie, 1996) 등에 의해 판정에 영향을 받기도 하였다. 따라서 본 연구에서 12.3%를 차지한 선수와 심판의 상호작용 요소가 무엇에 기인하는 것인지에 대하여 후속 연구를 통해 그 실체를 밝혀내고 이를 통해 심판 판정의 공정성을 더욱 높여야 할 것이다.

둘째, 피겨 스케이팅 연기영역 평가에서 적정수준의 일반화가능도를 확보할 수 있는 효율적인 측정 조건과 관련하여 심판 수가 많을수록 일반화가능도 계수가 높아졌다. 그러나 현재의 심판 수인 9명보다 많은 수의 심판이 참여하는 것은 효율성을 떨어뜨릴 수 있음을 고려하여 9명 이하에서의 일반화가능도 계수의 변화를 확인하였고, 0.9 이상의 일반화가능도 계수를 유지하기 위해서는 최소 2명 이상의 심판이 필요함을 확인하였다. 물론 평가 상황에서 요구되는 최소한의 일반화가능도 계수는 연구자가 결정하는 것이며, 0.8 이상의 일반화가능도 계수를 나타내는 경우에도 적정수준인 것으로 판단하는 경우도 있다(Mehrens & Lehmann, 1973). 따라서 2명 이하의 심판인 경우에도 적정수준의 일반화가능도 계수를 확보하는 것은 측정학적 관점에서 가능하다. 이러한 측정학적 측면의 결론은 현재 본 연구에 참여한 심판들의 판정이 측정학적으로는 안정적인 신뢰도를 확

보하고 있다는 것을 보여주는 것이다. 그러나 실제 대회 상황에서 심판의 수를 이와 같이 극단적으로 줄일 수는 없다. 왜냐하면 ISU(2014)에서는 시합이 공식적으로 그 결과를 인정받기 위해서는 Technical Panel과 Judge로 구성되는 심판진이 모두 최소 5명 이상이고, 이중 Judge는 최소 3명 이상을 규정하고 있기 때문이다. 반대로 심판진들 중에 Judge의 수가 9명을 넘을 때는 추첨을 통해 심판을 배정하고 있으며, 추첨을 할 심판의 선정에 있어서는 참가선수의 국적을 고려하여 추첨할 심판들을 선정한다. 사실 심판들의 판정은 자신과 선수의 국적에 따라 영향을 받을 수 있기 때문에(Ansorge & Scheer, 1986), 참가선수와 동일한 국적을 가진 심판들을 배제하는 것이 합리적인 선택일 수 있다. 그러나 피겨 스케이팅 강국에서 우수한 심판들이 배출되고 있는 것을 고려하면 이 방법은 심판수급/선정과 안정적인 대회운영에 부정적인 영향을 줄 수 있어서, ISU는 참가선수와 동일한 국적의 심판들이 심판 패널에 포함하는 정책을 선택하고 있다. 하지만 본 연구의 결과에서 제시하는 효율적인 측정 조건과 관련하여 0.9 이상의 일반화가능도 계수를 유지하기 위해서는 최소 2명 이상의 심판이 필요함을 확인되었다. 이는 현재 ISU가 경기를 공식적으로 인정하는 대회 규정인 5명 이상의 심판진 구성 원칙은 더 작은 숫자의 심판진 구성으로도 측정학적으로는 심판의 신뢰로운 채점이 가능함을 보여주고 있어, 대회운영의 효율성을 보다 더 높이는 측면에서 최소 심판수의 규정 개정도 가능하다.

물론 통계적 분석만을 고려하여 극단적으로 심판 수를 줄이는 것은 판정의 신뢰성에 또 다른 영향을 줄 수 있기 때문에 신중을 기해야 한다. ISU(2014)는 선수 연기 판정의 공정성과 신뢰성을 높이기 위해, 국제 대회의 쇼트프로그램의 경우 9명의 심판 판정 중 최고·최저점을 준 2명의 판정을 삭제하고 나머지 7명의 판정을 평균하여 점수를 산정하는 절삭 평균(trimmed mean) 방식을 적용하고 있다. 절삭 평균 산출방법은 극단적인 점수를 제거할 수 있는 장점이 있기 때문에(Looney, 1997), 피겨 스케이팅처럼 선수들의 연기가 종료된 후, 관찰자인 심판이 판정을 하는 체조, 다이빙, 리듬체조처럼 관찰자 오차가 존재할 수 있는 종목들에서 심판판정의 오류를 최소화하기 위해 사용되고 있다. 하지만 이러

한 방법만으로 관찰자 오차로서 존재하는 심판판정의 오류를 모두 예방할 수 있는 것은 아니다. 왜냐하면 이미 다수의 선행연구들(Ansorge & Scheer, 1986; Ansorge et al., 1978; Findlay & Ste-Marie, 2004; Ste-Marie, 1996)을 통해 현실에서 심판의 판정은 다양한 요인들에 영향을 받고 있는 것이 밝혀지고 있기 때문이다. 따라서 실제 스포츠 상황에서 심판의 수를 조절할 때에는 측정학적인 판단과 동시에 실제 경기 상황에 대한 전문가들의 논의도 있어야 할 것이다.

셋째, 일부 심판들은 다른 심판들과 달리 판정이 관대하거나 엄격하였으며, 구성요소 세부요소들에서도 그 경향이 동일하지 않았다. 특정 심판의 평균 점수가 다른 심판들에 비해 최대 0.34점의 차이를 보였으며, 특정 세부요소의 평균은 다른 세부요소에 비해 최대 0.39점의 차이를 보였다. 이러한 차이는 소수점 둘째자리까지 채점결과를 발표하는 피겨 스케이팅 종목에서 간과할 수 없는 차이이다. ISU(2014)는 심판의 질 관리를 위해 모든 심판들이 심판자격 취득 및 갱신을 위한 정기적인 세미나를 참석하여 수료하도록 하고 있다. 그리고 매 대회마다 심판들이 당일의 심판 판정의 적절성을 확인하는 Review Meeting 개최·운영을 의무화하고 있으며, 문제되는 판정이 발견될 때는 해당 심판을 징계한다. 하지만 이러한 노력에도 불구하고, 본 연구에서 일부 심판들의 판정은 동일한 경향을 나타내지 않고 있으며, 이는 일부 심판들의 판정의 엄격성이 다르게 나타난 Kim(2003)의 연구결과와도 동일하다.

분석결과를 심판과 세부요소 관점에서 구체적으로 살펴보면, 심판 P9는 심판 P2, P3, P4, P6과 통계적으로 유의미한 차이가 있는 판정을 한 것으로 나타났고, 심판 P5는 다른 심판들에 비해 판정의 표준편차가 크게 나타났는데 이러한 경우 P5와 P9 심판의 판정에 대한 논의가 다시 필요하다고 판단할 수 있다.

이 연구에서 분석한 심판들이 판정은 동간·비율척도(Table 2)에 의한 판정이었다. 따라서 두 심판(P5, P9)의 척도에 의한 판정이 다른 심판들과 달리 표준편차가 크고 평균이 통계적으로 유의미하게 낮은 것은 P5와 P9 심판의 척도를 통한 판정의 경향이 다른 심판들과 동일하지 않은 것을 의미한다. 물론 이러한 결과가 두 심판의 판정이 문제점을 가지고 있다는 것을 의미하지

않는다. 왜냐하면 나머지 7명의 판정이 타당하지 않을 수도 있기 때문이다. 이러한 점에 대해 해당 전문가들의 심도 있는 논의가 필요한 시점이다.

심판들 판정의 표준편차가 다른 요소들에 비해 상대적으로 컸던 구성요소의 안무/구성 요소는 판정의 문제점으로 지적될 수 있다.

선행연구들(Ahn, 2014; Cho & Chun, 1997)에서 피겨 스케이팅 종목 쇼트 프로그램의 측정오차는 프리 프로그램보다 크고, 선수들의 점수는 프리 프로그램보다 쇼트 프로그램에서 더 낮게 나타나기도 하였다. 특히 쇼트 프로그램 중에 본 연구의 분석이 이루어진 구성 점수는 의상과 음악 및 예술적인 표현과 관련되고 있어서, 일반화가능도 이론에서 주목하고 있는 관찰자 오차가 발생할 수 있을 것이라고 쉽게 예측이 가능하다.

ISU(2014)의 심판 판정 규정에서 안무/구성 요소는 공간, 형식, 구성, 통일, 분배, 악상에 따른 표현 등의 원칙에 따라서 모든 종류의 동작을 의도적으로 잘 다듬어 독창적으로 배열하여 만드는 것으로 정의하고 있다. 안무/구성 요소는 음악적 해석/타이밍 요소와 함께 구성요소에서 예술적 측면이 강한 평가요소이다. 예술적 측면이 강한 요소라는 의미는 이를 해석하는 심판의 주관에 그 판정의 영향이 크다는 것으로, 실제로 안무/구성 요소 및 음악적 해석/타이밍 요소와 관련된 심판의 판정은 선수의 연기에 적용된 안무나 음악을 이해하고 해석할 수 있는 심판의 능력에 영향을 받는다(Kim, 2013). 다시 말해, Zitzewitz(2014)가 지적하였듯이, 피겨 스케이팅에서 심판의 판정은 심판의 문화적 배경에 따라 영향을 받을 수 있는데, 이는 고전 및 현대 음악·뮤지컬·발레·무용 등 문화전반에 관한 심판의 이해 능력에 심판 판정이 영향을 받을 수 있다는 것을 의미하는 것이다. 따라서 동일한 선수의 연기에서 안무/구성 요소와 음악적 해석/타이밍 요소의 평가, 즉 예술적 평가는 심판들마다 상이하게 나타나기도 한다. 예를 들어, 국제 피겨 스케이팅 대회 심판 판정을 연구한 Ahn(2014)의 연구에서는 본 연구 결과와는 다르게 심판 판정에 있어 음악적 해석/타이밍 요소의 표준편차가 가장 컸다. 이는 심판들의 예술적 평가가 일관되지 않고 있음을 보여주는 사례라고 할 수 있다. 결과적으로 이러한 상황들을 통해 판정의 엄격성과 공정성이 훼손될 수 있음을 예상할 수 있다.

물론 첫 번째 논의에서 서술했듯이 본 연구에 참여한 심판들의 판정은 신뢰도 측면에서 높은 신뢰도 계수를 보여주었지만, 심판들 간의 판정의 경향은 동일하지 않았다. 이는 채점자내 신뢰도의 문제보다 현 상황에서는 채점자간 신뢰도의 문제가 보다 큰 문제일 수 있음을 보여준다. 다시 말해, 심판의 주관적 판단과 관련한 미적 요소 즉, 구성요소의 채점은 측정의 오차 측면에서 심판 내에서 뿐만 아니라, 심판 간에 크게 발생할 수 있음을 짐작하게 한다. ISU는 심판들이 동일한 판정의 엄격성을 유지하기 위한 세부요소 평가준거를 상세화하고 구체적인 심판 교육과 실습을 실천해 왔다. 하지만 본 연구의 결과는 이러한 노력이 심판의 주관적 판정의 위험을 완전히 제거하지는 못하고 있음을 보여주고 있다.

그렇다면 심판 판정 개선을 위한 대안은 무엇이 될 수 있는가? 본 연구 결과를 바탕으로 심판의 문화적 감수성을 높이는 제도적인 보완책 마련이 요청된다. 본 연구에서 표준편차가 가장 컸던 구성요소의 안무/구성과 예술적 요인인 음악적 해석/타이밍 요소의 해석은 심판의 문화적 감수성 능력에 영향을 크게 받는데, 이는 심판들이 각기 다른 문화적 배경과 경험을 갖고 있는 것에 기인한다. 지금까지 심판의 문화적 감수성 함양은 심판 개인의 책임 하에 개발해야 하는 영역으로서 여겨졌다. 따라서 심판들마다 문화적 감수성 능력은 편차가 존재하였고, 이는 심판 판정에 영향을 준다. 따라서 Ahn(2014)은 심판의 이러한 판정 경향을 고려하여 선수와 안무가가 피겨 연기를 위한 음악을 선정할 때, 지역의 특수성을 반영한 난해한 곡보다는 보다 더 대중적이고 폭넓게 공감할 수 있는 음악을 선정하는 것이 점수 획득에 유리함을 지적하였다. 물론 심판들의 문화적 감수성 능력의 편차는 문화권을 넘어 동일한 국적을 가진 심판들 간에도 발생하고 있다. 예를 들어, Kim(2013)이 피겨 스케이팅 종목 국내 및 국제심판들과 심층면담을 통해 제시한 자료에는 실제 국제 대회에서 일본 선수가 일본 전통 음악에 기반 한 연기에서 서구 심판들은 같은 아시아권인 한국 심판들보다 점수를 낮게 판정하는 상황이 발생하기도 했으며, 같은 국적의 심판이 같은 음악을 두고도 다른 의견을 표현하는 경우도 있었음을 보고하였다. 이러한 사례들은 심판의 문화적 감수성 함양을 위한 구체적인 심판 교육이 제도적 차원에서 접근되어야 함을 보여

준다. 다양한 문화적 배경을 갖는 심판들을 위해 문화적 감수성을 높이는 노력들이 요구된다.

요약하면, 이 연구에서 심판들은 채점자내에서 판정의 높은 신뢰도를 보여주었다. 그러나 전체 심판들 중에 일부는 그 경향이 다르게 나타나는 문제가 노출되었다. 따라서 심판 판정의 오차 범위를 최소화할 수 있도록 평가준거의 세부요소를 상세화하고 이를 적용하는 구체적인 심판 교육과 실습(Raymond et al., 1991; Tittle, 1982)은 지속적으로 실천되어야 하며, 특히 심판들의 문화적 감성을 높이는 체계적인 교육이 요구된다.

결론

연구결과에 근거하여 결론을 내리면 다음과 같다.

첫째, 심판들이 채점한 피겨 스케이팅 쇼트 프로그램 판정의 신뢰도는 높은 수준이었으며, 검사점수에 가장 영향을 미치는 것은 선수요인이었다.

둘째, 피겨 스케이팅 연기영역 평가의 효율적인 측정 조건과 관련하여 심판 수는 최소 2명 이상, 8명 이하에서도 적절한 수준의 평가가 가능하다.

셋째, 일부 심판들은 보다 엄격하거나 관대하게 평가하였으며, 심판의 일부 구성요소에 대한 평가와 관련하여 일관되지 않은 평가 경향이 나타났다. 이러한 결과는 판정과 관련하여 심판 재교육이 필요하다는 것을 보여준다.

본 연구는 일반화가능도 이론과 통계분석을 통해 심판의 관찰자 오차가 존재할 수 있는 피겨 스케이팅 종목에서 심판 판정의 신뢰도와 오차요인을 탐색하고, 심판 개인 및 연기 평가 세부영역의 평가 경향을 분석하여 심판 평가 개선을 위한 구체적인 대안을 제시한 초기의 연구라는 점에서 의의가 있다. 피겨스케이팅이나 리듬체조 종목처럼 채점자 오차가 존재하는 종목에서 일반화가능도 이론의 활용은 채점자 요인에 대한 신뢰성 연구를 기반으로 스포츠 현장에 유의미한 환류정보를 줄 수 있는 잇점이 있다(Lee & Yang, 2016). 이러한 분석방법은 대회 후 심판 재교육을 위한 다양한 자료를 제공할 수 있고, 이를 활용하면 해당 종목에서 심판에 의한 판정이 더욱 신뢰롭고 공정하게 이루어질 수 있도록 하는데 기여할 수 있어 앞으로 많은 스포츠 종목에서 그 활용을 기대하게 한다.

참고문헌

- Ahn, N. Y. (2006). Coaching of single figure skating by new judging system. *Journal of Coaching Development*, 8(1), 55-64.
- Ahn, N. Y. (2014). Correlation between technical score and program component score in figure skating. *Journal of Coaching Development*, 16(1), 199-205.
- An, J. K. (2014). A Study on the national university volleyball's referees's misjudgment triggered by negligence and resolution. *Journal of Korean Society for Sport Anthropology*, 9(2), 1-32.
- Ansonge, C. J., Scheer, J. K., Laub, J., & Howard, J. (1978). Bias in judging women's gymnastics induced by expectations of within-team. *Research Quarterly for Exercise and Sport*, 49(4), 399-405.
- Ansonge, C. J., & Scheer, J. K. (1986). *Test of a model scoring system for the self-regulation of international bias in gymnastic judging*. Paper presented at the Annual Meeting of the American Alliance for Health, Physical Education, Recreation and Dance in Cincinnati, Ohio.
- Ansonge, 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), 103-197.
- Brennan, R. L. (2001). *Generalizability theory*. NY: Springer-Verlag.
- Cha, J. H., Cha, Y. S., & Chang, K. R. (2015). Evaluation on the economic value of instant replay system in Korean Professional Baseball League using contingent evaluation method. *Korean Journal of Sport Management*, 20(3), 1-12.
- 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.
- Choi, J. S., & Choi, K. Y. (1999). An Analysis on the different views about JUDO Hantei Between referees and players. *The Korean Journal of Physical Education*, 38(1), 698-710.
- Cho, J. H., & Chun, M. K. (1997). Analysis of figure skating rating system. *J. Nat. Sci. Inst., Seoul Women's Univ.*, 9, 266-274.
- Crick, J. E., & Brennan, R. L. (1983). *Manual for GENOVA*:

- A generalized analysis of variance system* (ACT technical bulletin No. 43). Iowa City: American College Testing.
- Findlay, C., & Ste-Marie, M. (2004). A Reputation bias in figure skating judging. *Journal of Sport & Exercise Psychology*, 26, 154-166.
- Han, M. K., & Kim, S. H. (2009). A Study on the recognition of players towards the referee's decision in wheelchair basketball. *Journal of Rehabilitation Research*, 13(3), 75-92.
- ISU (2014). *Special regulations & technical rules: Single & pair skating and ice Dance 2014*. Switzerland, Lausanne: ISU.
- 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), 107-212.
- Kim, A. Y. (2013). *A Study on aesthetic experience in figure skating single program*. Master's degree Dissertation, Yonsei University.
- 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 Sport Science*, 13(2), 63-73.
- Kim, M. Y. (2003). Objectivity in the judgement of figure skating performance. *The Korean Journal of Physical Education*, 42(3), 745-755.
- Kim, N. H., & Jeon, J. W. (2014). A Study on the frequency of instant video replay during Taekwondo gyeorugi. *Journal of Korean Association of Physical Education and Sport for Girls and Women*, 28(3), 195-206.
- Kim, S. Y., & Kim, H. J. (2000). A Study of the influence of hockey judges' decision on athletes' and performance. *The Korean Journal of Physical Education*, 39(3), 110-118.
- Lee, J. H., Shin, S. H., & Jang S. H. (2014). An Analysis of research trend for figure skating revitalization. *Journal of Korean Society of Sport Policy*, 12(3), 167-178.
- Lee, T. K., & Yang, H. W. (2016). Analysis of error sources and estimation of reliability in peer review of forced connection method-sportscasting by applying generalizability Theory. *Korean Journal of Sport Science*, 27(2), 345-361.
- Loony, M. (1997). Objective measurement of figure skating performance. *Journal of Outcome Measurement*, 1, 143-163.
- Mehrens, W. A., & Lehmann, J. J. (1973). *Measurement and evaluation in educational and psychology*. NW: Holt, Rinehart, & Winston.
- Park, H. Y. (2003). The Analysis of cause Korean soccer referees' false judgment. *Korean Journal of Sport Psychology*, 14(3), 127-138.
- Park, J. J., & Park, S. S. (2012). The Understanding of referee's decision factor and improvement way in kundo competition. *The Journal of Korean Alliance of Martial Arts*, 14(3), 149-161.
- Raymond, M. J., Webb, L., & Houston, W. (1991). Correcting performance rating errors in oral examination. *Evaluation and the Health Professions*, 14, 100-222.
- SBS (2008.10.24.). A Little girl! Fascinate the world!
- Seo, K. H. (2011). Exploration on video decipher for fairness in basketball. *Philosophy of Movement: Journal of Korean Philosophic Society for Sport and Dance*, 19(2), 37-52.
- Shin, H. S., Park, J. H., Lee, S. Y., & Kim, J. H. (2010). Statistical analysis on 2008 World Championship figure skating data. *Journal of Natural Science*, 16, 79-87.
- Shin, Y. J., & Yoon, Y. J. (2013). A Study of stress sources and coping strategies in elite figure skating due to the introduction of new skating judging system. *Korean Journal of Sports Science*, 22(2), 483-499.
- Ste-Marie, D. (1996). International bias in gymnastic judging: Conscious or unconscious influences? *Perceptual and Motor Skill*, 83(3), 963-975.
- Tittle, C. K. (1982). Use of judgemental methods in item bias studies. *Handbook of methods for detecting test bias*, 31-63.
- Zitzewitz, E. (2014). Does transparency reduce favoritism and corruption? Evidence from the reform fo figure skating judging. *Journal of Sports Economics*, 15(1), 3-30.

<Appendix> Control Cards for GENOVA

```
GSTUDY      A X (C:P) DESIGN -- RANDOM MODEL
OPTIONS     RECORDS 2
EFFECT      * A  62 0
EFFECT      + P   9 0
EFFECT      + C:P 5 5
FORMAT      (5F5.0)
PROCESS
(Skip analysis data)
COMMENT
COMMENT     SET OF D STUDY CONTROL CARDS
DSTUDY      #1 -- A X (C:P) DESIGN -- C AND P RANDOM
DEFFECT     $ A
DEFFECT     P 9 8 7 6 5 4 3 2 1
DEFFECT     C:P 5
ENDDSTUDY
FINISH
```

피겨 스케이팅 심판 판정에 대한 분석

이태구 · 이한주 · 정재은 · 양희원(연세대학교)

이 연구의 목적은 일반화가능도 이론과 기술통계 분석을 통해 피겨 스케이팅 연기영역 판정의 신뢰도와 오차요인을 밝히고, 심판 판정 개선을 위한 대안을 제시하는 것이다. 일반화가능도 이론은 심판들이 선수들을 평가하는 측정상황에서 가능한 오차요인들의 영향력을 파악하고, 이를 바탕으로 신뢰도 계수를 향상시킬 수 있는 측정조건을 제시할 수 있다. 연구자료는 전국 피겨 스케이팅 대회 시니어 여자 싱글 1그룹에 참가한 62명의 선수들의 판정 기록 중, 쇼트 프로그램의 연기영역 판정 점수가 사용하였다. 9명의 심판들은 규정에 따라 연기영역의 세부요소들인 스케이팅 스킬, 트랜지션/연결 풋워크 & 동작, 연기/수행, 안무/구성과 해석/타이밍에 대하여 10.0 만점으로 수행 수준을 판단하여 각 영역 점수를 부여하였다. 연구자료에 대하여 일반화가능도 이론 및 기술통계, ANOVA 분석을 실시하였다. 연구 결과 첫째, G연구 결과 피겨 스케이팅 연기영역 평가는 일반화가능도 계수가 0.9 이상으로 높게 나와 안정적인 신뢰도를 확보하고 있었으며, 평가에 영향을 미치는 오차요인들 중 가장 영향력이 큰 것은 선수 요인이었다. 둘째, D연구 결과 심판을 2명까지 감소시켜도 0.9 이상의 안정적인 일반화가능도 계수를 보임을 확인하였다. 셋째, 기술통계 및 ANOVA 분석 결과 9명의 심판들은 대체로 동일한 평가 양상을 보였으나 소수의 심판들은 일부 연기영역 세부요소에서 차이를 보였다. 연구 결과를 근거로 현재 이루어지고 있는 피겨 스케이팅 연기영역 판정 방식의 신뢰성과 효율적인 경기 운영을 위한 판정 방식을 논의하였다. 또한 일관되지 않은 평가 경향을 드러내어 판정과 관련된 심판 교육의 필요성을 제기하였다.

주요어: 일반화가능도 이론, 피겨 스케이팅, 연기영역, 신뢰도, 오차요인, 판정경향