



Analysis of the Interrater Reliability and Agreement of Breaking Battles in the 2018 Buenos Aires Youth Olympics

Jin Su Suk^{1*} and Kyung Bae Jun²

¹Yongin University

²Hanyang University

Article Info

Received 2023.01.10.

Revised 2023.05.31.

Accepted 2023.06.13.

Correspondence*

Jin Su Suk

sucoi@yi.ac.kr

Key Words

Breaking battle,
Trivium value system,
Interrater reliability,
Interrater agreement

PURPOSE This study aimed to establish a standardized evaluation of the consistency and reliability of the results of the 2018 Buenos Aires Youth Olympics. **METHODS** The results of the preliminary and the final rounds of the individual Boys and Girls competition of the 2018 Buenos Aires Youth Olympics were collected. Data were organized using Microsoft Windows Excel 2021. Windows SPSS 26.0 was used for descriptive statistics. Intraclass correlation coefficient calculation was used to analyze the evaluation score. The winner/loser decision, and the evaluation interrater agreement was confirmed using Fleiss' kappa. **RESULTS** There was a difference in the scores given by the judges in both the preliminary and final rounds of both competitions. A low correlation was observed between the decided winner of the judges, and the point variance for each item. **CONCLUSIONS** In the trivium value system, differences in the evaluation strategies of the judges may exist. Thus, improvements to standardize evaluation should be implemented. In addition, this scoring system considers various factors in the evaluation process, and the scoring system is complicated. Therefore, there may be a need to revise the evaluation system and find the optimal number of judges to create a better evaluation.

서론

다수 심판의 관찰과 주관에 의해 대상자의 수행능력을 심사하거나 경기결과를 결정하는 방법은 무용, 체육, 스포츠 분야에서 전통적으로 사용된 방법이다(Cho, 1999; Kang, 2011). 왜냐하면, 난이도·완성도·정확성과 같은 기술적 측면뿐 아니라 음악성·창의성·예술성과 같은 미적 측면까지 종합적으로 고려한 심사를 진행할 수 있기 때문이다(Choi, 2011; Kim & Jeon, 2018; Lee & Kim, 2020). 이러한 심사 방법은 심사에 참여하는 심판이 가진 이론적(theoretical), 전문적(expertise) 지식과 경험적(empirical) 지식을 기반으로 종목 또는 각 분야의 특성으로 고려하여 사전에 설정한 기준을 이용하여 대상자를 심사하게 된다(Anshell, 1990; Choi, 2011; Kim & Lee, 2021; Premelč et al., 2019). 이에 심판의 주관에 의한 심사를 진행하는 분야에서는 체계적이고 객관성 있는 심사를 위하여 세부 항목별 기준을 설정하고 그에 따른 점수 척도를 구성하여 현장에서 적극적으로 활용하고 있다(Cho &

Choi, 2015; Choi & Kang, 2004; Lee et al., 2017).

'브레이킹(Breaking)'은 비보잉(B Boying)으로 알려져 있는 브레이크 댄스(Break Dance)로 힙합(Hip Hop) 또는 스트리트 댄스(Street Dance)의 한 분야인 문화적 요소로 받아들여져 왔다. 하지만, 2018년 부에노스 아이레스 유스올림픽과 2024년 파리 올림픽에 정식종목으로 채택되며 하나의 댄스스포츠로서 메가 스포츠 이벤트(Mega Sport event) 정식종목으로 인정받게 되었다(Cha, 2021; Moon & Kim, 2021; Park et al., 2021). 브레이킹의 스포츠화 과정에서 경기 방식은 브레이킹 발생부터 사용되고 있는 '배틀(battle)'을 그대로 반영하여 모든 대회에 사용 중에 있다(Lee, 2022; Park et al., 2021). 배틀은 춤으로 실력을 겨루어 승자와 패자를 결정하는 경연의 형태로 대상자들은 정해진 시간 또는 횟수 내에 서로 번갈아 가며 자신의 기술을 선보이게 된다(KFD, 2022b). 배틀을 심사하기 위하여 기존에는 다수의 심판이 참여하여 다수결에 의한 판정을 하거나 5점 또는 10점 만점을 기준으로 점수를 부여해 승자를 선정하는 방법을 이용해 왔다(Jun & Suk, 2022; Lee, 2022; Park et al., 2021). 하지만 기존의 배틀 심사 방법은 명확하지 않은 심사 기준과 점수체계, 그리고 결과의 객관성과 관련한 문제 제기가 지속적으로 이루어져 왔으며 이는 브레이킹

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

의 스포츠 종목으로 편입을 반대하는 주요 의견으로 제시되기도 하였다(Park et al., 2021).

이러한 문제를 해결하고자 WDSF(World Dance Sport Federation : 세계댄스스포츠연맹)에서는 브레이킹 심사제도 구축 및 체계화 과정에서 브레이킹의 특성을 고려한 심사를 위한 '트리폴드(ThreeFold)'와 '트리비움 밸류 시스템(Trivium Value System)'을 개발하여 보급하고, 현장 심사에 사용 중에 있다(Fahr et al., 2018; Jun & Suk, 2022). 이 두 가지 브레이킹 심사 방법은 모두 전통적으로 브레이킹에서 중요하게 생각하는 세 가지 영역(Domain)인 'Body : Physical(신체)', 'Mind : Artistic(예술)', 'Soul : Interpretive(표현, 해석)'를 기본으로 하위 심사요인을 구분 및 계층화하여 만든 심사 방법으로 WDSF와 함께 비보이 스톰(B Boy Storm), 디제이 레니게이드(DJ Renegade)에 의해 2017년 개발되었다(Fahr et al., 2018; Jun & Suk, 2022; WDSF, 2017). 대회의 규모나 운영방법에 따라 적용되는 심사 방법에 대한 규정이 없어 트리폴드와 트리비움 밸류 시스템 중 주최 측에서 자유롭게 선택하여 사용하며 일반적으로 국내 대회 같은 소규모 대회에서는 트리폴드를 사용하며 세계선수권, 올림픽과 같은 대규모 국제대회에서는 트리비움 밸류 시스템을 브레이킹 배틀 심사에 사용하고 있다(And8. dance, 2022b; Fahr et al., 2018; WDSF, 2019; WDSF, 2022a).

트리폴드와 트리비움 밸류 시스템의 세 영역의 심사 준거는 다시 각각 2개의 영역으로 구분된다. 'Physical'은 'Technique'과 'Variety'로 'Artistic'은 'Creativity'와 'Personality', 'Interpretive'는 'Performativity'와 'Musicality'로 세분화하고 있다(WDSF, 2017; USA Dance, 2022). 이에 따른 심사 준거 및 하위 영역은 두 가지 심사 방법이 동일하나 채점 방법에 차이가 있다. 트리폴드는 'Physical', 'Artistic', 'Interpretive'의 세 가지 심사 준거에 각각 33.3%의 비중을 두어 감점, 가점 요소를 종합해 점수를 부여하고, 트리비움 밸류 시스템은 'Physical' 'Technique' 20%와 'Variety' 13.333%로 'Artistic'은 'Creativity' 20%, 'Personality' 13.333%로 'Interpretive'는 'Performativity' 20%와 'Musicality' 13.333%로 총 6가지 항목을 구분하여 채점한다. 이때, 트리비움 밸류 시스템은 감점과 가점 요인에 대하여 별도의 채점 방법을 이용해 보다 세분화 된 심사와 점수를 부여하게 된다(Fahr et al., 2018; USA Dance, 2022). 트리폴드와 트리비움 밸류 시스템을 이용한 브레이킹 배틀의 심사에는 결과의 객관성 확보를 위하여 5명의 심판이 참여하도록 하고 있으며 대회별로 두 가지 방법 중 선정된 심사 방법을 이용하여 심사를 진행한다. 이때, 5명의 심판은 매 배틀 별로 대상자의 춤을 보고 사전에 선정된 심사 기준을 이용하여 주관적 심사를 진행하게 된다(Fahr et al., 2018).

브레이킹 심사와 같이 다수 심판의 주관에 의해 심사를 진행하는 무용, 체육, 스포츠 분야 현장과 연구 분야에서 결과의 신뢰도와 심사의견의 일치도는 지속적인 관심의 대상이 되어왔다. 왜냐하면, 모든 심판이 대상자의 수행능력에 대해 유사한 점수를 부여할 수 없으며, 심사 점수에서 심판 간 차이가 없다고 해도 같은 의견을 가지고 동일한 항목에 점수를 부여했는지는 알 수 없기 때문이다(Kim & Jeon, 2018; Choi, 2011; Lee & Kim, 2020). 이에 주관적 심사를 활용하는 분야에서는 적극적으로 심사결과를 분석하는 연구가 수행되고 있다. Choi는 2010년과 2011년 현대 무용 실기능력 평가 자료를 분석한 연구를 통해 동일한 기준을 가지고 심사하여도 심판 개인 특성 차이, 엄격성 차이 등으로 인해 심사결과가 다르게 나타남

을 확인하였고, Kim & Jeong(2006)와 Kim & Jeon(2018)의 연구에서는 댄스스포츠 경기결과 결정을 위해 사용 중인 스케이팅 시스템(Skating System : 절대과반수 우위판정법)의 심사결과를 분석한 결과 심사의견 일치 수준에 차이가 있으며 심판 간 신뢰도가 낮은 경향을 보이기 때문에 현재 사용 중인 순위 결정 방법의 개선이 필요함을 보고하였다. 또한, Lee & Kim(2020)과 Kim & Lee(2021)은 댄스스포츠 라틴 심사결과에서 심판의 등급과 경력, 심사 위치와 심사 시기 등 심판의 개인적 특성과 심사 조건 등에 따라 결과의 일치도가 다르게 나타나는 경향이 있어 심판의 차이를 줄일 수 있는 교육 프로그램의 개선을 제안하였다. 브레이킹의 경우 Jun & Suk(2022)의 연구에서 항목과 대상자에 따라 심판이 부여하는 점수에 차이가 있으며 이로 인해 심판 간 신뢰도와 일치도가 낮게 나타나 대회 결과에 대한 신뢰성 확보에 어려움이 있음을 보고하였다. 이렇듯 여러 선행 연구에서는 심판의 주관에 의존하는 심사 상황에서 동일한 심사 기준과 방법을 활용하여도 심사의 질적 결과인 심사의견과 그에 따른 양적 결과인 점수에 심판 간 차이가 나타날 수 있어 지속적인 현장 자료에 대한 검증과 이를 기반으로 한 심판 교육 및 제도 개편에 대한 필요성을 이야기하고 있다.

브레이킹 심사에 활용 중인 트리폴드와 트리비움 밸류 시스템은 비교적 최근에 개발된 심사 방법으로 심판의 주관에 의존하여 심사를 진행하기 때문에 선행연구에서 나타난 문제점들이 유사하게 나타날 수 있어 실제 심사결과를 분석한 정량적 검증이 필요하다. 하지만 국내·외에서 실제 브레이킹 심사 자료를 활용한 연구는 트리폴드를 이용해 진행된 국내 대회 자료를 분석한 Jun & Suk(2022)의 연구로 제한적이며 트리비움 밸류 시스템을 통해 진행된 심사결과에서 '다수 심판이 동일한 점수를 부여했는가'를 의미하는 심판 간 신뢰도와 '동일한 심사의견을 가지고 심사를 하였는가'를 의미하는 심사의견의 일치도를 분석하는 연구는 아직 이루어지지 않고 있는 실정이다. 따라서, 본 연구는 트리비움 밸류 시스템을 이용하여 실시된 최초의 메가스포츠 이벤트인 2018년 부에노스 아이레스 유스올림픽 심사결과와 심판 간 신뢰도 및 일치도를 분석하여 보다 나은 브레이킹 배틀 심사제도 구축을 위한 기초자료를 현장에 제공하고자 한다.

연구방법

연구자료

본 연구에서는 2018년 부에노스 아이레스 유스올림픽에서 열린 브레이킹 배틀 중 개인전인 B Boy(남자), B Girl(여자)의 예선과 결선 결과를 분석자료로 선정하였다. B Boy, B Girl 모두 예선은 라운드 로빈(Round Robin) 방식으로 진행되었으며 결선은 8강 토너먼트로 운영되었다. 이때 모든 경기는 5명의 심판이 참여하여 심사를 진행하였다. 전체 경기 수 및 연구자료의 특성은 <Table 1>과 같다.

Table 1. Characteristic subjects

Variables		Total battles	Total rounds	Judge
B Boy	Preliminaries	18	36	5
	Final	8	32	
B Girl	Preliminaries	18	36	5
	Final	8	32	

브레이킹 경기 및 심사 특성

1. 경기 구성

개인전과 단체전 구분 없이 모든 브레이킹 경기장은 <Figure 1>과 같이 구성된다. 브레이킹 배틀이 이루어지는 'Dance Floor(Battle Ground)'는 전체 경기장의 중앙에 위치하며 압력 성형 목재 메이소나이트(Masonite)를 사용한다. 'Dance Floor'의 규격은 최소 '가로 6m x 세로 6m의 사각형'이나 최소 '지름 6m의 원형' 등으로 경기장 특성 또는 대회 규모에 따라 다양하게 구성할 수 있으며 'Dance Floor'의 바닥 양쪽에 각각 'Blue'와 'Red'를 표시한다. 경기장의 중앙에는 MC(Master of Ceremonies)가 위치하여 전체적인 경기의 진행을 담당한다. 심사에 참여하는 심판의 수는 대회 규모에 따라 차이가 있으나 올림픽을 비롯한 국제대회에서는 헤드 저지(Head Judge) 2명을 포함한 총 7명의 심판이 참여한다. 이 중 5명의 심판은 'Dance Floor' 뒤편에 위치하며 배틀을 심사하고, 2명의 헤드 저지는 경기장 측면에 위치하여 올바른 절차에 의해 심사가 진행되도록 관리한다. 모든 심사결과는 실시간으로 전산 처리되어 저장되며

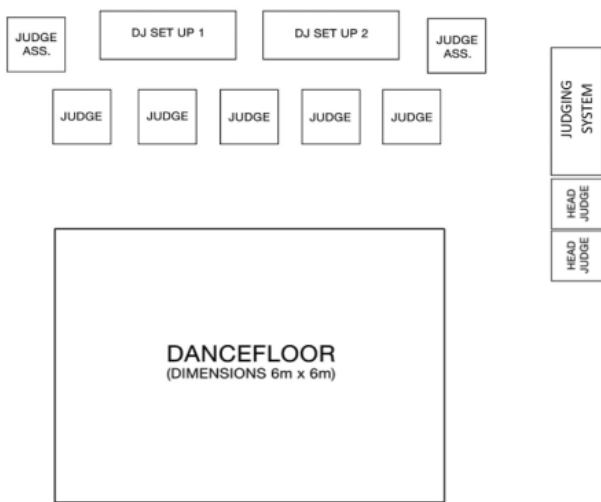


Fig. 1. Stage and dance floor of breaking (Fahr et al., 2018)

배틀이 종료되면 즉시 현장에서 공개되는데 심사결과 수집 및 공개를 위한 전산시스템(And8.dance)과 심사결과 확인을 위한 전광판이 경기장의 규모와 특성을 고려하여 설치된다. 배틀을 위한 음악은 경기장의 뒤편 또는 측면에 두 대의 DJ 장비(DJ Set up)를 준비하여 DJ(Disc Jockey)가 배틀의 흐름이 끊기지 않도록 음악을 준비하여 제공하게 된다(Fahr et al., 2018).

현재 브레이킹의 경우 배틀에 사용되는 음악의 종류와 속도에 대한 규정은 없으며, 경기 주최 측에서 음악을 선택하도록 하고 있다(WDSF, 2021a). 이에 주최 측에서는 전체적인 대회 진행을 고려하여 배틀에 필요한 음악을 준비하며, 각 라운드별로 다르게 제공되는 음악에 맞추어 선수들은 자신의 기량을 선보이게 된다(Fahr et al., 2018).

2. 경기 운영방법

2018 부에노스 아이레스 유스올림픽 브레이킹 시드배정 및 경기 운영방법의 예시는 <Figure 2>와 같다. 모든 예선전은 라운드 로빈으로 운영되었다. 라운드 로빈은 모든 참가자가 차례로 서로 대결하는 방법으로 B Boy와 B Girl 각각 12명의 참가자가 돌아가며 1:1 배틀을 진행하였다. 이때 각각의 배틀은 2라운드로 실시 되었으며 2라운드의 모든 배틀에서 보다 많은 심판이 승자로 판정된 결과를 종합하여 예선 순위를 결정하였다.

본선은 예선 결과 8위까지의 선수를 대상으로 시드를 배정하여 토너먼트를 구성해 실시되었다. 토너먼트 구성에서 선수는 각각 'Blue'와 'Red'로 구분되며 시드배정 과정 중 상대적으로 예선전 순위가 낮은 선수가 'Red', 순위가 높은 선수가 'Blue'를 배정받게 된다. 토너먼트의 배틀 진행은 'Red' 선수가 먼저 시작하며 'Red' 선수의 댄스가 종료되면 'Blue' 선수가 댄스를 선보이게 된다. 배틀에서 라운드 별로 각각 2회의 기회가 선수들에게 주어지며 1회에 선수는 40초~1분 이내에 자신의 댄스를 선보여야 한다. 본선 토너먼트의 각각 배틀은 4라운드로 진행되며 승자 결정은 4라운드에서 보다 많은 심판이 승자로 선정한 결과를 종합하여 선정하였다(Fahr et al., 2018, WDSF, 2021a).

3. 트리비움 밸류 시스템(Trivium Value System)

트리비움 밸류 시스템의 심사 요소와 각각의 계층 구조는 <Table 2>와 같다. 심사 요소는 크게 'Physical', 'Artistic', 'Interpretative' 구분되며 각각 33.333%의 배점을 가진다.

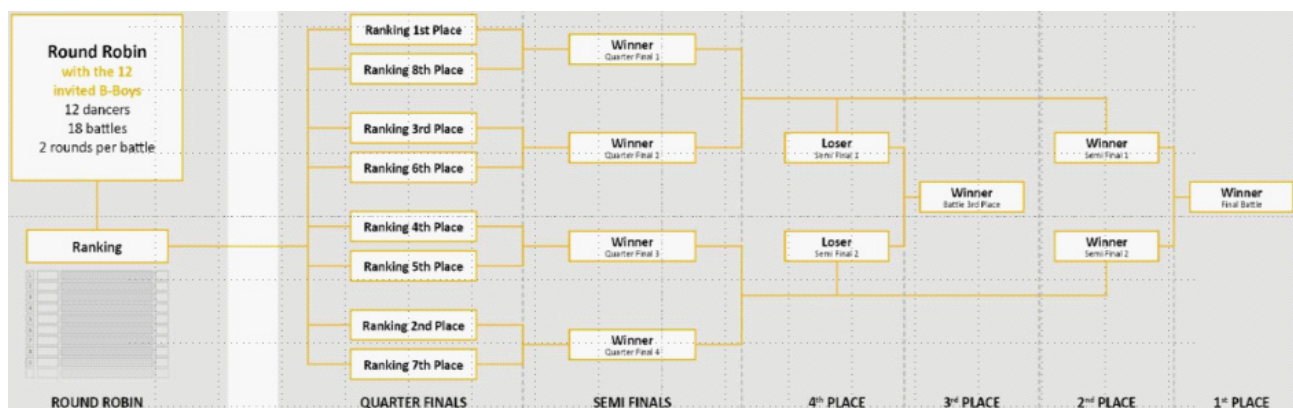


Fig. 2. Breaking match progress and seeding of 2018 Buenos Aires Youth Olympic Games (Fahr et al., 2018)

‘Physical’은 신체적 능력에 대한 요인으로 20%의 ‘Technique’과 13.333%의 ‘Variety’로 구분되며 신체적 능력을 통해 발휘되는 기술과 경기력 외에 움직임의 범위와 다양한 동작의 수행을 심사한다. ‘Interpretative’는 ‘표현적 능력’ 또는 ‘해석적 능력’으로 정의되며 제공되는 음악에 맞추어 배틀에 대한 구성과 참가자가 자신이 표현하고자 하는 것을 춤으로 표현하고 있는지를 ‘Performativity’ 20%와 ‘Musicality’ 13.333%로 구분하여 심사하게 된다. ‘Artistic’은 예술적 능력을 심사하기 위하여 기초기술을 기반으로 얼마나 독창성 있는 춤을 선보이는데 대한 ‘Creativity’ 20%, 무대 장악력과 개인의 개성을 충분히 선보일 수 있는 ‘Personality’ 13.333%를 종합하여 심사한다(USA Dance, 2022; WDSF, 2019).

브레이킹 배틀 심사는 전산화 되어 있는 트리비움 밸류 시스템을 이용하여 진행된다. 배틀에 심사 참여하는 심판들은 모두 트리비움 밸류 시스템이 설치되어있는 개별 태블릿(Tablet Computer)을 배정받아 심사를 진행한다. 트리비움 밸류 시스템의 화면의 구성은 <Figure 3>과 같다. 트리비움 밸류 시스템은 선수들이 배틀에서 보여주는 기량을 ‘Physical’, ‘Artistic’, ‘Interpretative’의 3가지 대분류 기준에 대해 ‘Technique’, ‘Variety’, ‘Creativity’, ‘Personality’, ‘Performativity’, ‘Musicality’의 6가지 중분류 항목을 배정하여 ‘Blue’와 ‘Red’ 선수를 심사할 수 있도록 구성되어 있으며, 각각 세부 항목에 대한 크로스 패이더(Cross Fader)를 움직여 보다 뛰어난 경기력을 보여준 선수에 대해 점수를 부여하게 된다. 이때, 심판들은 각각의 심사 요소를 기준으로 관찰을 통한 주관적 심사를 통해 참가자를 심사하며 트리비움 밸류 시스템을 이용해 ‘Technique’, ‘Performativity’, ‘Creativity’는 0~20.0%의 범위에서 ‘Variety’, ‘Musicality’, ‘Personality’는 0~13.333%의 범위 내에서 자유롭게

점수를 부여할 수 있다(USA Dance, 2022; WDSF, 2019).

크로스 패이더를 이용한 항목별 점수 부여 외에 숏컷(Shortcut) 버튼을 통해 추가적인 가점과 감점을 할 수도 있다. 숏컷 버튼은 숏컷 점수 버튼(Shortcut-Score Buttons), 숏컷 패널티 버튼(Shortcut-Penalty Buttons)으로 구분되며, 각 버튼을 사용할 때마다 정해진 점수 척도에 따라 해당 버튼에 배정된 항목별 점수의 가점 또는 가점이 이루어진다(Fahr et al., 2018; WDSF, 2019).

숏컷 점수 버튼 특성은 <Table 3>과 같다. 숏컷 점수 버튼은 총 4가지인 ‘Execution’, ‘Form’, ‘Confidence’, ‘Spontaneity’ 구성되며 각 항목은 가점을 위한 ‘+’ 및 감점을 위한 ‘-’의 점수 버튼을 사용할 수 있다. 숏컷 점수 버튼 중 ‘Execution’, ‘Form’ 버튼을 1회 누를 때마다 ‘Technique’와 ‘Performativity’ 점수에서 ‘2.00%’씩 가감된다. 또한, ‘Confidence’ 버튼을 1회 사용할 때마다 ‘Performativity’와 ‘Personality’에서 ‘1.66%’씩, ‘Spontaneity’는 ‘Performativity’, ‘Creativity’의 점수에서 ‘2.00%’씩 감점 또는 가점이 발생해 점수에

Table 2. Hierarchy and Value of Trivium Judging system

Categories and value		Criteria hierarchy
Physical (body) (33.3%)	Technique (20.0%)	<ul style="list-style-type: none"> ▶ Aptitude & athleticism ▶ Dexterity-fine motor control ▶ Pushing the body to the limit ▶ Balance ▶ Strength & endurance ▶ Flexibility ▶ Dynamics ▶ Spatial control
	Variety (13.333%)	<ul style="list-style-type: none"> ▶ Dimension ▶ Empathetic approach
Interpretative (soul) (33.3%)	Performativity (20.0%)	<ul style="list-style-type: none"> ▶ Composition & narrative ▶ The element of surprise ▶ Emotional engagement ▶ Authenticity
	Musicality (13.333%)	<ul style="list-style-type: none"> ▶ Coherence ▶ Accenting ▶ Syncopation ▶ Texture ▶ Phrasing
Artistic (mind) (33.3%)	Creativity (20.0%)	<ul style="list-style-type: none"> ▶ Progression of the fundamentals of foundation
	Personality (13.333%)	<ul style="list-style-type: none"> ▶ Stage presence & charisma ▶ Individuality & character

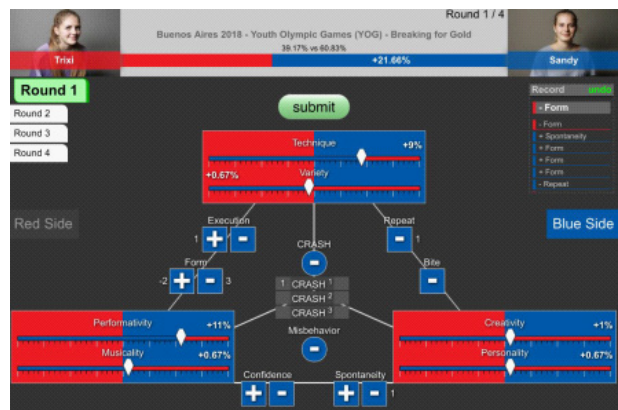


Fig. 3. Trivium value system interface (WDSF, 2022a)

Table 3. Characteristic of shortcut-score buttons and additional or deduction points

Buttons	Influence criteria	Grades per push
Execution	Technique, performativity	2.00%
Form	Technique, performativity	2.00%
Confidence	Performativity, personality	1.66%
Spontaneity	Performativity, creativity	2.00%

Table 4. Characteristic of shortcut-penalty buttons and deduction points

Buttons	Influence Criteria	Grades per Push
Repeat	Variety, Creativity	1.66%
Bite	Variety, Creativity, Personality	2.34%
Crash	All Faders	1 Push = slip : 2.5%
		2 Push = Crash : 5.00%
		3 Push = wipe-out : elimination
Misbehavior	Performativity, Creativity, Personality	1 Push = mild : 2.66%
		2 Push = moderate : 5.34%
		3 Push = severe : 8.00%

반영되게 된다(Fahr et al., 2018; WDSF, 2019).

슛컷 패널티 버튼의 특성은 <Table 4>와 같다. 슛컷 패널티 버튼은 두 개의 심플(Simple Penalty), 두 개의 콤플렉스(Complex Penalty) 버튼 총 4개로 구성되며 슛컷 패널티 버튼의 경우 슛컷 점수 버튼과 다르게 가산점은 없이 ‘-’버튼을 통해 감점에만 사용된다. 심플 버튼을 사용한 감점의 경우 선수가 배틀 중 자신이 이미 수행한 기술을 반복 수행하면 심판은 ‘Repeat’ 버튼을 누르게 되며 1회 누를 때마다 ‘Variety’, ‘Creativity’의 항목점수에서 ‘1.66%’가 감점된다. 또한, 타 선수의 독창적 기술을 모방한 경우 ‘Bite’ 버튼을 사용하며 ‘Variety’, ‘Creativity’, ‘Personality’의 점수에서 ‘2.34%’의 감점이 이루어진다. 콤플렉스 버튼의 ‘Crush’는 선수가 춤 또는 기술을 선보이는 중 기술의 실패와 관련된 요인으로 1회 사용할 때마다 모든 항목별 점수에서 ‘2.5%’가 감점되며, 심판이 3회 이상 사용한 경우 해당 선수는 실격 처리된다. ‘Misbehavior’ 버튼은 선수의 춤 또는 기술에서 나타나는 실수보다는 잘못된 행동이나 적절하지 못한 태도가 발생했을 경우 사용한다. 배틀 중 대상자가 인종차별이나 외설적, 폭력적 언행과 같이 브레이킹에 대한 인식에 악영향을 미칠 수 있는 행동을 하는 경우 사용하며 버튼 사용횟수에 따라 ‘Performativity’, ‘Creativity’, ‘Personality’에서 정해진 점수 척도에 따라 감점이 이루어진다. 이때, 3인 이상 심사위원들의 ‘Misbehavior’ 사용할 경우 선수는 실격되며 배틀에 참가하는 선수가 지나치게 잘못된 행위를 하는 경우 ‘Misbehavior’ 버튼 사용 여부와 상관없이 즉시 실격시킬 수도 있다(Fahr et al., 2018; WDSF, 2019).

4. 심사결과 특성

모든 브레이킹 심사결과와의 전산처리하는 ‘And8.dance’에서 담당하고 있다(And8.dance, 2022a). 트리비움 벨류 시스템을 통해 진행된 브레이킹 심사결과는 실시간으로 전산을 통해 저장되며 매 배틀이 종료된 이후 현장 전광판에 승·패 결과와 항목별 세부 심사결과가 공개된다. 모든 경기가 종료되면 저장된 최종 결과는 ‘And8.dance’ 홈페이지에 게시된다. 최종 브레이킹 배틀 결과는 크게 심판이 라운드 별 승·패 여부에 대해 선택한 다수결 결과와 트리비움 벨류 시스템을 통해 부여한 항목별 점수로 구분된다. 심판의 다수결 결과의 경우 개별 심판이 라운드 별 승자로 선택한 결과와 5명 심판이 승자로 선택한 결과가 함께 제시된다. 항목별 점수의 경우 ‘Blue’와 ‘Red’의 선수에 대해 심판이 부여한 ‘Technique’, ‘Variety’,

‘Creativity’, ‘Personality’, ‘Performativity’, ‘Musicality’의 세부 점수를 각각 확인할 수 있다. 이와 더불어 점수 영향을 주는 슛컷 버튼의 사용 내역을 심판별로 확인할 수 있다(And8.dance, 2022a). 최종 결과의 예시는 <Figure 4>와 같다.

자료처리

2018년 부에노스 아이레스 유스올림픽 브레이킹 심사결과와의 심판 간 신뢰도 및 일치도 분석을 위해 ‘https://and8.dance/’에서 ‘예선’과 ‘결선’ 결과를 다운로드 받아 Window Excel 2021을 활용하여 통계 방법에 맞게 정리하였다. 브레이킹 심사에서 심판은 ‘Blue’와 ‘Red’에 각각 점수를 부여하기 때문에 라운드 별로 선수에게 심판이 부여한 항목별 점수를 구분하였다. 또한, 모든 배틀의 매 라운드별 심판이 선택한 승자와 패자의 선정 결과와 함께 심사에 사용한 감점, 가점 내용인 슛컷 내용을 함께 정리하였다. 이때, ‘Blue’를 선택한 경우는 ‘1’, ‘Red’를 선택한 경우는 ‘2’의 숫자를 부여하였다. 정리된 모든 자료는 Window SPSS 26.0을 이용하여 자료처리를 실시하였다.

1. 항목별 심판 간 신뢰도

항목별로 5명의 심판이 부여한 점수에 대한 심판 간 점수의 차이를 확인하기 위해 심사점수의 기술통계를 통해 점수의 평균, 표준편차를 산출하였다. 이후 심판 간 신뢰도 확인을 위해 ICC(Intraclass Correlation Coefficient : 유목내상관계수)를 산출하였다. 브레이킹 배틀의 경우 임의로 선정된 대상자(Random)에 대해 고정되어(Fixed)있는 5명의 심판이 심사를 진행하게 된다. 따라서 ICC 산출 분석을 위한 모형은 ‘2원 혼합모형(Two Way Mixed)’ 효과 분석의 신뢰도 유형은 ‘일관성(Consistency)’으로 선정하여 선정된 심판들의 점수에 대한 결과를 확인하였다(Kang, 2011). 이때, 산출된 ICC는 한명의 심판의 심사했을 때의 단일 ICC(single)와 다수 심판들의 평균값을 이용한 평균 ICC(average) 중 심판 간 신뢰도 확인을 위한 본 연구의 목적에 맞추어 다수 심판의 평균 ICC를 이용하여 연구결과를 확인하였으며, 심동적 영역의 심판 간 신뢰도 인정수준인 .80을 기준으로 검증하였다(Oh & Lee, 2002).

2. 다수결 결과의 일치도

브레이킹 배틀 심사결과와 각각 라운드 별로 부여된 심사점수 신



Fig. 4. Breaking battle report of using trivium (And8.dance, 2022a)

외도 외에 5명 심판이 결정한 승자와 패자 선정 여부와 함께 항목별로 심판 심사의견에 심판 간 일치도를 확인하기 위하여 플레이스 카파 (Fleiss' Kappa)를 산출하였다. 플레이스 카파는 '동일 측정 대상들에 대하여 3명 이상의 심판들이 범주형으로 유사한 결과의 일치 정도'를 분석할 수 있는 통계량이다. 산출된 플레이스 카파계수는 Landis & Koch(1977)의 기준에 따라 심판 간 일치도를 확인하였다. Landis & Koch(1977)의 플레이스 카파 해석 기준은 <Table 5>와 같다.

Table 5. Landis & Koch's ranges and agreement of kappa

Kappa statistic	Strength of agreement
<0.00	Poor
0.00~0.20	Slight
0.21~0.40	Fair
0.41~0.60	Moderate
0.61~0.80	Substantial
0.81~1.00	Almost perfect

연구결과

본 연구의 목적은 2018년 부에노스 아이레스 유스올림픽 브레이킹 심사결과와 심판 간 신뢰도 및 심사의견의 일치도를 확인하는데 있다. 이에 따라 연구결과는 B Boy, B Girl의 예선과 결선의 결과를 각각 나누어 제시하였다. 심판 간 신뢰도 분석결과표는 'Blue'와 'Red' 선수에게 부여한 심판의 항목별 점수를 함께 확인할 수 있도록 구성하였으며, 심판 간 일치도 분석결과와 승·패 선정, 항목별 심사의견과 감점 및 가점을 위해 'Blue'와 'Red' 선수에게 각각 사용한 숫컷 버튼의 일치도 분석결과를 함께 제시하였다.

B Boy 분석결과

1. B Boy 예선 심판 간 신뢰도

B Boy 예선 전체 결과 심판별 점수의 기술통계 결과와 심판 간 신뢰도 분석결과는 <Table 6>과 같다. B Boy Blue의 항목별 ICC는 Technique 0.722, Variety 0.674, Performativity 0.755, Musicality 0.487, Creativity 0.348, Personality 0.295였으며,

Table 6. Judge score result and intraclass correlation coefficient of B Boy preliminaries

Variables	B Boy blue				ICC	Variables	B Boy red				ICC
	Judge	M	SD	ICC			Judge	M	SD	ICC	
Technique (N=36)	J1	2.47	2.70	0.722	Technique (N=36)	J1	1.00	1.63	0.732		
	J2	1.64	1.62			J2	0.86	1.79			
	J3	2.02	2.11			J3	0.44	0.97			
	J4	1.45	1.41			J4	0.23	0.78			
	J5	1.52	1.29			J5	0.83	1.28			
Variety (N=36)	J1	1.04	1.24	0.647	Variety (N=36)	J1	0.29	0.73	0.074		
	J2	0.33	0.73			J2	0.04	0.22			
	J3	0.55	0.89			J3	0.30	0.57			
	J4	0.49	0.71			J4	0.06	0.33			
	J5	0.95	0.84			J5	0.56	0.83			
Performativity (N=36)	J1	0.68	1.41	0.755	Performativity (N=36)	J1	0.21	0.60	0.702		
	J2	1.64	1.76			J2	1.08	1.74			
	J3	2.33	2.23			J3	0.44	1.18			
	J4	1.37	1.15			J4	0.52	0.98			
	J5	1.53	1.45			J5	1.01	1.46			
Musicality (N=36)	J1	0.82	1.20	0.487	Musicality (N=36)	J1	0.03	0.18	0.192		
	J2	0.54	1.56			J2	0.04	0.22			
	J3	0.73	1.05			J3	0.22	0.67			
	J4	0.36	0.56			J4	0.14	0.47			
	J5	1.15	0.96			J5	0.43	0.72			
Creativity (N=36)	J1	1.56	2.01	0.348	Creativity (N=36)	J1	1.03	1.78	0.662		
	J2	1.13	1.52			J2	0.33	1.06			
	J3	0.21	0.58			J3	0.06	0.33			
	J4	0.97	1.00			J4	0.02	0.13			
	J5	1.44	1.31			J5	0.83	1.27			
Personality (N=36)	J1	0.46	0.80	0.295	Personality (N=36)	J1	0.38	0.73	0.579		
	J2	0.11	0.39			J2	0.23	0.61			
	J3	0.09	0.40			J3	0.00	0.00			
	J4	0.63	0.73			J4	0.25	0.58			
	J5	1.00	1.00			J5	0.67	0.86			

B Boy Red의 항목별 ICC는 Technique 0.732, Variety 0.074, Performativity 0.702, Musicality 0.192, Creativity 0.662, Personality 0.579로 나타났다. ICC 산출을 통한 심판 간 신뢰도 분석결과 Blue와 Red에게 부여한 모든 항목의 점수에서 나타나는 ICC가 항목별로 차이가 있었다. 또한, 산출된 항목별 ICC는 모든 항목에서 인정수준에 미치지 못하는 결과를 보여 낮은 심판 간 신뢰도 경향을 보이는 것으로 나타났다.

2. B Boy 예선 심판 간 일치도

B Boy 예선 승·패 여부에 대한 심판 간 의사결정 및 항목별 심사결과와 일치도 분석결과는 <Table 7>과 같다. 예선전 승·패 여부에 대한 일치도는 0.469로 '보통 수준 일치(moderate)'로 나타났다. 항목별 심사의견의 일치도의 경우 Technique은 0.257로 '어느 정도 일치(fair)', Variety는 0.102로 '약간의 일치(slight)', Performativity는 0.221로 '어느 정도 일치'를 보이는 것으로 분석되어 심판 간 심사 의견에 차이가 있는 것으로 나타났다. 이외에 Musicality, Creativity, Personality 항목의 심사의견의 경우 통계적으로 유의한 모형이 나타나지 않아 항목별 심사의견이 일치하지 않는 것을 확인할 수 있었다.

B Boy 예선의 심판이 사용한 슛컷 버튼의 일치도 분석결과는 <Table 8>과 같다. B Boy Blue의 경우 Execution, Form, Confidence, Spontaneity, Repeat의 총 5개의 슛컷 버튼을 심판이 심사에 사용하였다. 이때 심판 간 의견의 일치도 분석결과 Execution, Repeat는 통계적으로 일치도 모형이 나타나지 않았으며 통계적으로 유의한 일치도 모형이 나타난 Form은 -0.125, Confidence은 -0.226, Spontaneity은 -0.224로 음수의 결과를 보여 일치도 측면에서 의미 없는 결과를 보였다. B Boy Red의 분석결과 8개의 슛컷 버튼 중 Execution, Form, Confidence, Spontaneity, Repeat, Crash 1 - Slip의 6개 버튼을 사용한 것

Table 7. Result of decision interrater agreement of B Boy preliminaries

Variables	kappa	p	95% CI
Total decision	0.469	< 0.001	0.465~0.472
Technique	0.257	< 0.001	0.255~0.260
Variety	0.102	0.010	0.099~0.104
Performativity	0.221	< 0.001	0.218~0.223
Musicality	0.070	0.096	0.067~0.073
Creativity	0.022	0.577	0.020~0.025
Personality	-0.014	0.722	-0.016~-0.011

Table 8. Result of shortcut-penalty decision interrater agreement of B Boy preliminaries

B Boy Blue				B Boy Red			
Variables	kappa	p	95% CI	Variables	kappa	p	95% CI
Execution	-0.004	0.930	-0.007~-0.001	Execution	0.073	0.112	0.070~0.076
Form	-0.125	0.013	-0.128~-0.121	Form	-0.061	0.208	-0.064~-0.058
Confidence	-0.226	< 0.001	-0.229~-0.223	Confidence	-0.205	< 0.001	-0.209~-0.202
Spontaneity	-0.224	< 0.001	-0.227~-0.220	Spontaneity	-0.197	< 0.001	-0.201~-0.194
Repeat	-0.006	0.916	-0.009~-0.002	Repeat	-0.011	0.831	-0.015~-0.008
				Crash 1 - Slip	0.361	< 0.001	0.357~0.364

으로 나타났다. B Boy Red의 슛컷 버튼 사용의 일치도 분석결과 Execution, Form, Repeat의 경우 통계적으로 유의한 모형이 나타나지 않았다. 통계적으로 유의한 일치도 모형이 나타난 결과 중 Crash 1 - Slip는 0.361로 '어느 정도 일치'를 보였으나 Confidence와 Spontaneity는 각각 -0.205와 -0.197의 음수의 결과로 의미 없는 일치도 결과를 나타내었다.

B Boy 예선의 심판 간 신뢰도 및 일치도 분석결과를 종합해서 살펴보면 항목별로 산출된 ICC에 차이가 있는 것으로 나타났으며 항목별 점수의 심판 간 신뢰도는 모든 항목에서 낮은 신뢰도를 보였다. 특히, B Boy Red에 부여된 Variety와 Musicality의 경우 다른 항목의 심사결과에 비교하였을 때 심판 간 신뢰도가 매우 낮은 것으로 분석되었다. 승·패 여부에 대한 심판 간 의사결정 및 항목별 심사결과와 일치도 분석결과 역시 항목별로 심사 의견에 차이가 있는 것으로 나타났다. 또한, 의사결정 결과뿐 아니라 항목별 심판의 의견과 슛컷 버튼 사용 결과 역시 낮은 수준의 일치도를 보이는 것을 연구결과를 통해 확인할 수 있었다.

3. B Boy 결선 심판 간 신뢰도

B Boy 결선의 기술통계 결과와 심판 간 신뢰도 분석결과는 <Table 9>와 같다. 항목별 점수의 심판 간 신뢰도 분석결과 B Boy Blue의 경우 Technique 0.687, Variety 0.578, Performativity 0.620, Musicality 0.329, Creativity 0.628, Personality 0.544로 모든 항목에서 인정수준에 미치지 못하는 결과를 보였다. B Boy Red의 항목별 심판 간 신뢰도를 위해 산출한 ICC는 Technique 0.735, Variety 0.769, Performativity 0.574, Musicality 0.353, Creativity 0.797, Personality 0.377으로 B Boy Red의 항목별 점수 역시 모든 항목에서 인정수준에 미치지 못하는 낮은 심판 간 신뢰도를 보이는 것으로 나타났다.

4. B Boy 결선 심판 간 일치도

B Boy 결선의 승·패 여부에 대한 심판 간 의사결정과 항목별 심판 의견의 일치도 분석결과는 <Table 10>과 같다. 승·패 여부 결정의 일치도는 0.580으로 '보통 수준 일치'였다. 항목별 심판의 심사의견의 일치도는 0.312의 Technique과 0.256의 Variety, 0.248의 Creativity는 '어느 정도 일치'였고, 0.144의 Performativity는 '약간의 일치'로 분석되었으며 Musicality와 Personality는 통계적으로 유의한 심판 간 일치도 모형이 나타나지 않았다. 심판의 승·패 여부에 대한 심판 간 의사결정의 항목별 일치도 분석결과 심판 간 일치도 경향에 차이가 있었으며 일부 항목에서는 심판 간 일치도가 나타나지 않았다.

B Boy 결선에서 심판의 슛컷 버튼 사용에서 나타나는 심판 간 일치도 분석결과는 <Table 11>과 같다. B Boy 결선에서는 총 8

Table 9. Judge score result and intraclass correlation coefficient of B Boy final

Variables	B Boy blue				Variables	B Boy red			
	Judge	M	SD	ICC		Judge	M	SD	ICC
Technique (N=32)	J1	1.41	2.01	0.687	Technique (N=32)	J1	1.29	1.94	0.735
	J2	1.73	2.05			J2	1.31	2.00	
	J3	1.28	1.65			J3	0.46	0.88	
	J4	1.40	1.30			J4	0.65	1.17	
	J5	1.64	1.96			J5	0.89	1.16	
Variety (N=32)	J1	0.47	0.78	0.578	Variety (N=32)	J1	0.23	0.78	0.769
	J2	0.58	0.88			J2	0.67	0.88	
	J3	0.88	0.97			J3	0.30	0.66	
	J4	0.45	0.71			J4	0.50	0.86	
	J5	0.86	1.08			J5	0.45	0.60	
Performativity (N=32)	J1	1.58	2.09	0.620	Performativity (N=32)	J1	0.83	1.32	0.574
	J2	0.62	1.45			J2	0.25	0.67	
	J3	1.16	1.90			J3	0.51	0.86	
	J4	1.18	1.17			J4	0.66	1.13	
	J5	1.36	1.79			J5	0.70	0.93	
Musicality (N=32)	J1	0.18	0.52	0.329	Musicality (N=32)	J1	0.30	0.82	0.353
	J2	0.29	0.79			J2	0.30	0.62	
	J3	1.03	1.45			J3	0.63	1.00	
	J4	0.35	0.60			J4	0.05	0.28	
	J5	1.28	1.41			J5	0.57	0.84	
Creativity (N=32)	J1	0.72	1.19	0.628	Creativity (N=32)	J1	0.36	1.06	0.797
	J2	0.68	1.36			J2	0.43	0.72	
	J3	0.81	1.12			J3	0.38	0.77	
	J4	0.61	1.08			J4	0.65	1.08	
	J5	1.66	1.90			J5	0.73	1.09	
Personality (N=32)	J1	0.48	0.86	0.544	Personality (N=32)	J1	0.04	0.24	0.377
	J2	0.16	0.63			J2	0.01	0.07	
	J3	0.23	0.45			J3	0.00	0.00	
	J4	0.42	0.69			J4	0.33	0.68	
	J5	1.02	1.18			J5	0.38	0.54	

Table 10. Result of Decision Interrater Agreement of B Boy Final

Variables	kappa	p	95% CI
Total decision	0.580	< 0.001	0.576~0.583
Technique	0.312	< 0.001	0.309~0.315
Variety	0.256	< 0.001	0.253~0.258
Performativity	0.144	< 0.001	0.142~0.147
Musicality	-0.013	0.743	-0.016~-0.011
Creativity	0.248	< 0.001	0.245~0.250
Personality	0.075	0.079	0.072~0.077

개의 슛컷 버튼 중 B Boy Blue와 Red 모두 Execution, Form, Confidence, Spontaneity, Repeat, Crash 1 - Slip의 6개 슛컷 버튼이 사용한 것으로 나타났다.

B Boy Blue의 슛컷 버튼 사용 일치도 분석결과 Execution과 Crash 1 - Slip은 통계적으로 유의한 일치도 모형이 나타나지 않았으며 Repeat는 0.176로 '약간의 일치' 수준을 보였고, Form은 -0.207,

Confidence는 -0.194, Spontaneity는 -0.212의 음수의 결과를 보여 일치도 측면에서 의미 없는 결과를 보였다. B Boy Red의 슛컷 버튼 사용의 심판 간 일치도 분석결과 Execution은 통계적으로 유의한 일치도 모형이 나타나지 않았으며 Repeat, Crash 1 - Slip는 각각 0.189, 0.145로 '약간의 일치'였으며, Form은 -0.200, Confidence는 -0.241, Spontaneity는 -0.212의 음수 일치도가 산출되었다.

B Boy 결선의 분석결과를 종합하여 살펴보았을 때 심판 간 신뢰도의 경우 B Boy Blue와 Red 심사결과에서 모두 낮은 신뢰도를 보이고 있었으며, 항목별 심판 간 신뢰도 경향 역시 차이가 있는 것으로 분석되었다. 또한, B Boy Blue와 Red의 연구결과에서 심판 간 신뢰도 경향에도 차이가 있는 것으로 나타나 항목별, 대상별로 부여한 심판 간 점수에 차이가 있는 것을 연구결과를 통해 확인하였다.

심사점수 이외에 승·패 결정 여부와 항목별, 슛컷 버튼 사용에 나타나는 심판 간 일치도 분석결과 분석된 모든 항목에서 심판 간 일치도 경향에 차이가 있고, 일치도가 나타난 경우에도 매우 낮은 수준에 일치도를 보이고 있었다. 이러한 결과는 B Boy 예선과 결선에서 모두 공통적으로 나타나고 있어 심사의견과 그에 따라 부여되는 심사 점수에 심판 간에 많은 차이가 존재하는 것을 확인할 수 있었다.

B Girl 분석결과

1. B Girl 예선 심판 간 신뢰도

B Girl 예선 심판별 점수의 기술통계 결과와 심판 간 점수의 신뢰도 분석결과는 <Table 12>와 같다. B Girl Blue의 항목별 분석결과를 살펴보면 Technique 0.731, Variety 0.590, Performativity 0.686, Musicality 0.622, Creativity 0.491, Personality 0.381로

항목별 심판 간 신뢰도 경향에 차이가 있었으며, 산출된 ICC는 모두 인정수준에 미치지 못하는 결과를 보였다. B Girl Red 예선 점수의 심판 간 신뢰도 분석결과 Technique 0.833, Variety 0.544, Performativity 0.495, Musicality 0.167, Creativity 0.494, Personality 0.630으로 분석되어 Technique을 제외한 모든 항목의 심판 간 신뢰도는 낮은 것으로 나타났으며 항목별 심판 간 신뢰도 경향에 차이가 있었다.

Table 11. Result of shortcut-penalty decision interrater agreement of B Boy final

B Boy blue				B Boy red			
Variables	kappa	<i>p</i>	95% CI	Variables	kappa	<i>p</i>	95% CI
Execution	-0.030	0.574	-0.033~-0.027	Execution	-0.033	0.482	-0.036~-0.030
Form	-0.207	< 0.001	-0.210~-0.203	Form	-0.200	< 0.001	-0.204~-0.197
Confidence	-0.194	< 0.001	-0.197~-0.191	Confidence	-0.241	< 0.001	-0.244~-0.237
Spontaneity	-0.212	< 0.001	-0.216~-0.209	Spontaneity	-0.212	< 0.001	-0.216~-0.209
Repeat	0.176	0.002	0.172~0.179	Repeat	0.189	0.001	0.186~0.193
Crash 1 - Slip	-0.013	0.821	-0.016~-0.009	Crash 1 - Slip	0.145	0.010	0.141~0.148

Table 12. Judge score result and intraclass correlation coefficient of B Girl preliminaries

B Girl blue					B Girl red				
Variables	Judge	<i>M</i>	<i>SD</i>	<i>ICC</i>	Variables	Judge	<i>M</i>	<i>SD</i>	<i>ICC</i>
Technique (<i>N</i> =36)	J1	2.94	2.65	0.731	Technique (<i>N</i> =36)	J1	0.58	1.35	0.833
	J2	1.95	2.26			J2	0.57	1.47	
	J3	2.61	1.56			J3	0.28	0.97	
	J4	2.19	1.57			J4	0.18	0.78	
	J5	2.27	1.26			J5	0.19	0.58	
Variety (<i>N</i> =36)	J1	1.06	1.32	0.590	Variety (<i>N</i> =36)	J1	0.34	0.69	0.544
	J2	0.17	0.49			J2	0.17	0.55	
	J3	0.40	0.64			J3	0.07	0.31	
	J4	0.75	0.82			J4	0.18	0.54	
	J5	1.15	0.78			J5	0.17	0.41	
Performativity (<i>N</i> =36)	J1	1.15	1.48	0.686	Performativity (<i>N</i> =36)	J1	0.11	0.46	0.495
	J2	2.57	2.56			J2	0.97	2.69	
	J3	3.26	1.83			J3	0.39	1.50	
	J4	1.34	1.10			J4	0.14	0.58	
	J5	2.00	1.16			J5	0.21	0.53	
Musicality (<i>N</i> =36)	J1	0.70	1.08	0.622	Musicality (<i>N</i> =36)	J1	0.30	0.62	0.167
	J2	0.37	0.79			J2	0.04	0.22	
	J3	0.52	1.03			J3	0.00	0.00	
	J4	0.83	0.86			J4	0.24	0.67	
	J5	1.30	0.86			J5	0.20	0.43	
Creativity (<i>N</i> =36)	J1	1.40	2.07	0.491	Creativity (<i>N</i> =36)	J1	0.31	0.82	0.494
	J2	1.03	1.49			J2	0.74	1.34	
	J3	0.72	1.08			J3	0.33	0.89	
	J4	1.12	1.19			J4	0.00	0.00	
	J5	1.79	1.15			J5	0.20	0.59	
Personality (<i>N</i> =36)	J1	0.90	1.01	0.381	Personality (<i>N</i> =36)	J1	0.23	0.52	0.630
	J2	0.18	0.47			J2	0.18	0.47	
	J3	0.22	0.50			J3	0.07	0.31	
	J4	0.75	0.90			J4	0.17	0.50	
	J5	0.86	0.95			J5	0.46	0.55	

2. B Girl 예선 심판 간 일치도

B Girl 예선 승·패 결정과 항목별 심사의견의 일치도 분석결과는 <Table 13>과 같다. 전체 라운드에서 5명의 심판이 내린 승·패에 대한 의사결정의 일치도는 0.468로 '보통 수준 일치'로 나타났다. 항목별 세부 심사의견의 심판 간 일치도는 Technique 0.280로 '어느 정도 일치'였으며, Variety와 Performativity는 각각 0.084, 0.083로 '약간의 일치'를 보였으나 나머지 항목인 Musicality, Creativity, Personality는 통계적으로 유의한 일치도 모형이 나타나지 않았다.

B Girl 예선의 심판 간 일치도 분석결과 승·패 결과 결정에 대한 심판 의견과 일부 항목의 세부 심사의견에서만 심사의견에 일치도가 있었으나 높지 않은 수준의 일치도 경향을 보이는 것으로 분석되었다.

B Girl 예선에서 사용된 숏컷 버튼 사용의 심판 간 일치도 분석결과는 <Table 14>와 같다. B Girl 예선에서 심판은 Blue 심사에 총 8개의 숏컷 버튼 중 Execution, Form, Confidence, Spontaneity, Repeat, Crash 1 - Slip, Misbehavior의 7개 버튼을 사용하였으며, Red 심사에서는 Execution, Form, Confidence, Spontaneity, Repeat, Crash 1 - Slip의 6개 버튼을 사용한 것으로 나타났다.

B Girl Blue 예선 숏컷 버튼 사용에서 심판 간 일치도 분석결과 Execution, Form, Crash 1 - Slip, Misbehavior 4개의 항목에서는 통계적으로 유의한 모형이 나타나지 않았다. 통계적으로 유의한 일치도 모형이 나타난 세 항목의 경우 Repeat는 0.099로 '약간의 일치'를 보였고, Confidence와 Spontaneity는 각각 -0.178, -0.159의 음수 결과를 보여 일치도 측면에서 의미 없는 결과를 보였다. B Girl Red 예선 숏컷 버튼 사용의 심판 간 일치도 분석결과

과 Execution, Form, Repeat, Crash 1 - Slip의 4가지 항목에서는 심판 간 일치도에 통계적으로 유의한 모형이 나타나지 않았다. Confidence와 Spontaneity는 통계적으로 유의한 모형은 나타나지 않으나 각각 -0.175, -0.099의 음수 결과를 보여 일치도 해석에는 의미 없는 결과를 보였다.

B Girl 예선의 분석결과를 종합적으로 살펴보면 심판 간 신뢰도는 Blue와 Red의 모든 항목별 점수에서 낮은 신뢰도를 보이고 있었으며 항목별 신뢰도 수준에 차이가 있는 것으로 나타났다. 또한, 심판 간 의견의 일치도 분석결과 역시 모든 전체 승·패 결정과 항목별 심사의견과 감점, 가점을 위해 사용한 숏컷 버튼 사용에서 낮은 일치도 수준을 보여 심판 간 심사의견에 차이가 있었다.

3. B Girl 결선 심판 간 신뢰도

B Girl 결선의 심판별 점수 기술통계 결과와 심판 간 신뢰도 분석결과는 <Table 15>와 같다. B Girl Blue의 항목별 분석결과 Technique 0.772, Variety 0.694, Performativity 0.864, Musicality 0.634, Creativity 0.729, Personality 0.790으로 Performativity를 제외한 나머지 항목의 ICC는 인정수준에 미치지 못하는 것으로 나타났으며 항목별 심판 간 신뢰도 경향에 차이가 있었다. B Girl Red의 항목별 심판 간 신뢰도 분석결과와 경우 Technique 0.681, Variety 0.428, Performativity 0.581, Musicality 0.554, Creativity 0.543, Personality 0.496으로 모든 항목에서 인정수준에 미치지 못하는 낮은 심판 간 신뢰도 경향을 보였다.

B Girl 결선의 분석결과 B Girl Blue와 Red 모두 항목 심판 간 신뢰도 경향에 차이가 있었으며, 대상자에게 부여한 점수의 심판 간 신뢰도 역시 Blue와 Red 간에 차이가 있었다. 분석결과 중 B Girl Blue Performativity 결과에서는 인정수준 보다 높은 ICC가 산출되었고, 산출된 ICC는 높은 수준의 심판 간 신뢰도를 나타내었지만, 이를 제외한 B Girl Blue와 Red의 모든 분석결과에서 낮은 심판 간 신뢰도 수준을 보이는 것을 연구결과를 통해 확인할 수 있었다.

4. B Girl 결선 심판 간 일치도

B Girl 결선의 승·패 결정과 항목별 심사의견의 일치도 분석결과는 <Table 16>과 같다. 승·패 여부에 대한 의사결정의 일치도 분석결과 0.373으로 '어느 정도 일치'로 나타났다. 항목별 심사의견의 심판 간 일치도 분석결과 Technique은 0.258로 승·패 여부와 유사한 '어느 정도 일치' 수준을 보였으나 나머지 항목의 경우 Variety 0.126, Performativity 0.155, Musicality 0.112, Creativity 0.192,

Table 13. Result of decision interrater agreement of B Girl preliminaries

Variables	kappa	p	95% CI
Total decision	0.468	< 0.001	0.465~0.471
Technique	0.280	< 0.001	0.277~0.282
Variety	0.084	0.039	0.081~0.086
Performativity	0.083	0.038	0.080~0.085
Musicality	0.072	0.080	0.070~0.075
Creativity	0.024	0.554	0.021~0.026
Personality	0.051	0.186	0.049~0.054

Table 14. Result of shortcut-penalty decision interrater agreement of B Girl preliminaries

B Girl blue				B Girl red			
Variables	kappa	p	95% CI	Variables	kappa	p	95% CI
Execution	-0.071	0.077	-0.073~-0.068	Execution	-0.059	0.162	-0.061~-0.056
Form	-0.026	0.573	-0.029~-0.023	Form	-0.054	0.177	-0.057~-0.052
Confidence	-0.178	< 0.001	-0.181~-0.175	Confidence	-0.175	< 0.001	-0.178~-0.172
Spontaneity	-0.159	0.001	-0.162~-0.156	Spontaneity	-0.099	0.047	-0.102~-0.096
Repeat	0.099	0.034	0.096~0.102	Repeat	-0.012	0.829	-0.015~-0.008
Crash 1 - Slip	-0.008	0.840	-0.011~-0.006	Crash 1 - Slip	-0.023	0.662	-0.027~-0.020
Misbehavior	-0.006	0.916	-0.009~-0.002				

Table 15. Judge score result and intraclass correlation coefficient of B Girl final

Variables	B Girl blue				Variables	B Girl red			
	Judge	<i>M</i>	<i>SD</i>	<i>ICC</i>		Judge	<i>M</i>	<i>SD</i>	<i>ICC</i>
Technique (<i>N</i> =32)	J1	2.35	2.57	0.772	Technique (<i>N</i> =32)	J1	0.73	1.29	0.681
	J2	2.06	2.35			J2	0.55	1.14	
	J3	1.01	1.92			J3	1.71	1.88	
	J4	1.15	1.26			J4	0.84	1.09	
	J5	1.02	1.39			J5	0.94	1.02	
Variety (<i>N</i> =32)	J1	0.54	1.12	0.694	Variety (<i>N</i> =32)	J1	0.16	0.44	0.428
	J2	0.74	1.09			J2	0.28	0.66	
	J3	0.47	0.82			J3	0.61	0.90	
	J4	0.52	0.71			J4	0.12	0.38	
	J5	0.64	1.18			J5	1.14	1.10	
Performativity (<i>N</i> =32)	J1	1.83	2.01	0.864	Performativity (<i>N</i> =32)	J1	1.01	1.77	0.581
	J2	0.89	1.74			J2	0.38	0.81	
	J3	1.24	2.39			J3	2.14	2.47	
	J4	0.80	1.17			J4	0.89	1.03	
	J5	0.79	1.42			J5	1.18	1.12	
Musicality (<i>N</i> =32)	J1	0.47	1.00	0.634	Musicality (<i>N</i> =32)	J1	0.17	0.68	0.554
	J2	0.33	0.86			J2	0.93	0.91	
	J3	0.09	0.34			J3	0.59	0.90	
	J4	0.22	0.57			J4	0.66	0.69	
	J5	0.45	0.73			J5	0.83	0.80	
Creativity (<i>N</i> =32)	J1	0.75	1.41	0.729	Creativity (<i>N</i> =32)	J1	0.38	0.79	0.543
	J2	1.09	1.77			J2	0.33	1.00	
	J3	0.83	1.32			J3	0.73	1.01	
	J4	0.49	0.97			J4	0.49	0.75	
	J5	1.08	1.77			J5	1.57	1.69	
Personality (<i>N</i> =32)	J1	0.42	1.02	0.790	Personality (<i>N</i> =32)	J1	0.19	0.66	0.496
	J2	0.25	0.64			J2	0.40	0.73	
	J3	0.07	0.28			J3	0.12	0.34	
	J4	0.13	0.38			J4	0.59	0.64	
	J5	0.43	0.73			J5	0.72	0.76	

Table 16. Result of decision interrater agreement of B Girl final

Variables	kappa	<i>p</i>	95% CI
Total decision	0.373	< 0.001	0.370~0.377
Technique	0.258	< 0.001	0.255~0.261
Variety	0.126	0.001	0.124~0.129
Performativity	0.155	< 0.001	0.152~0.157
Musicality	0.112	0.006	0.110~0.115
Creativity	0.192	< 0.001	0.190~0.195
Personality	0.127	0.002	0.124~0.129

Personality 0.127로 분석되어 모두 '약간의 일치' 수준의 일치도 경향을 보였다. 승-패 결정, 항목별 심사의견 심판 간 일치도 분석결과 배틀의 승-패 의사결정 및 모든 항목의 세부 심사의견에서 낮은 일치도 경향을 보여 심사 의견에 차이가 있는 것으로 나타났다.

B Girl 결선의 슛컷 버튼 사용 심판 간 일치도 분석결과(〈Table 17〉)과 같다. 총 8개의 슛컷 버튼 중 Blue와 Red 심사에 모두 Execution,

Form, Confidence, Spontaneity, Repeat, Crash의 6개 버튼을 심판이 사용하였으며 B Girl Blue 심사에서는 Crash 버튼을 두 번 눌렀을 때 발생하는 감점인 Crash 2 - Crash가 사용되어 심사에 반영된 것으로 나타났다.

Blue 심사의 슛컷 버튼 사용에 대한 심판 간 일치도 분석결과 Execution, Form, Crash 1 - Slip, Crash 2 - Crash에서는 통계적으로 유의한 일치도가 나타나지 않았다. 통계적으로 유의한 일치도 모형을 보인 항목의 경우 Repeat만 0.142로 '약간의 일치'였으며 -0.181의 Confidence, -0.212의 Spontaneity는 모두 음수의 결과로 일치도 해석 측면에서는 의미 없는 결과가 나타났다. Red 심사에 사용된 슛컷 버튼 사용의 일치도의 경우 Execution과 Crash 1 - Slip은 통계적으로 유의한 일치도 모형이 나타나지 않았으며, 나머지 항목에서 Repeat만 0.096의 결과로 모두 '약간의 일치'로 낮은 수준의 일치도를 보이고 Form은 -0.158, Confidence는 -0.186, Spontaneity는 -0.153로 음수의 일치도가 산출되었다.

B Girl 결선의 분석결과를 종합해서 살펴보면 심판 간 신뢰도의 경우 B Girl Blue Performativity의 점수에서만 높은 수준의 심판 간 신뢰도를 나타내었고, 그 외의 항목과 대상자의 점수에서는 낮은

Table 17. Result of shortcut-penalty decision interrater agreement of B Girl final

B Girl blue				B Girl red			
Variables	kappa	p	95% CI	Variables	kappa	p	95% CI
Execution	-0.031	0.494	-0.034~-0.028	Execution	-0.024	0.595	-0.026~-0.021
Form	-0.001	0.989	-0.004~-0.003	Form	-0.158	0.002	-0.161~-0.155
Confidence	-0.181	0.001	-0.185~-0.178	Confidence	-0.186	0.001	-0.189~-0.183
Spontaneity	-0.212	< 0.001	-0.215~-0.208	Spontaneity	-0.153	0.004	-0.156~-0.149
Repeat	0.142	0.002	0.139~0.145	Repeat	0.096	0.046	0.093~0.099
Crash 1 - Slip	0.071	0.204	0.067~0.074	Crash 1 - Slip	-0.032	0.564	-0.036~-0.029
Crash 2 - Crash	-0.006	0.910	-0.010~-0.003				

수준의 심판 간 신뢰도를 보였다. 또한, 심판 간 일치도 역시 승-패 결정과 항목별 심사의견과 감점 및 가점을 위한 슛컷 버튼의 사용에서 모두 낮은 수준의 일치도 경향이 나타났다.

B Boy와 B Girl의 예선과 결선 분석결과를 통해 살펴볼 때 브레이킹 배틀에 참여한 심판이 부여하는 점수와 점수를 부여함에 있어 각각의 심판이 가지고 있는 심사의견에 차이가 있는 것을 연구결과를 통해 확인할 수 있었다.

논의

본 연구의 목적은 브레이킹 배틀 심사를 위해 개발된 트리비움 밸류 시스템을 이용하여 심사가 이루어진 2018년 부에노스 아이레스 유스올림픽 브레이킹 결과에 대한 심판 간 신뢰도 및 일치도를 분석하여 보다 나은 브레이킹 배틀 심사의 기초자료를 현장에 제공하는데 있다. 이에, 자세한 연구결과를 확인하기 위해 본 연구에서는 라운드 로빈으로 진행된 예선과 토너먼트로 진행된 결선, 그리고 남·녀로 구분된 심사결과를 세부적으로 분석하여 각각의 심판 간 신뢰도 및 일치도를 확인하였다.

연구결과 B Girl 결선 Blue의 Performativity 결과를 제외한 모든 B Boy, B Girl의 예선과 결선 결과에서 심판 간 신뢰도는 매우 낮았고, 심사 점수에서 나타나는 심판 간 신뢰도 경향은 항목별로 차이가 있었다. 이러한 결과는 B Boy와 B Girl의 Blue와 Red 선수에게 부여된 모든 점수에서 유사하게 나타났다. 또한, 승-패 결정과 6가지 세부 항목별 심사의견의 일치도, 심사에 사용한 감점과 가점 요인에 대한 슛컷 버튼 사용의 심판 간 의견의 일치도 역시 낮은 것으로 분석되었으며 일치도 경향 역시 항목과 대상자별로 차이가 있었다. 본 연구결과는 같은 대회 내에서 트리비움 밸류 시스템이라는 동일한 심사 도구를 이용하여 심사를 진행하여도 심사에 참여하는 5명 심판 간 심사의견이 다르며 그로 인해 예선과 결선, 선수의 성별, 배틀에 참여한 두 대상자에게 항목별로 부여하는 점수에 차이가 있는 것을 의미한다.

무용과 댄스스포츠 분야 선행연구에서도 본 연구와 유사하게 같은 심사 기준을 사용하더라도 심사결과에 심판 간 차이가 있음을 보고하고 있다. Jun & Suk(2022)의 연구에서는 트리폴드를 사용한 브레이킹 배틀 심사결과에서 5명 심판이 부여한 점수와 심사의견에 차이가 있으며 이로 인해 승-패 결과뿐 아니라 심사 세부 영역에서 모두 낮은 심판 간 신뢰도와 일치도 결과가 나타났다. Oh et al.(2001), Choi(2010, 2011)은 무용의 수행능력을 심사할 때 같은 심사 도구

와 기준을 사용하여도 심판 간 개인의 특성과 엄격성과 차이로 인해 부여하는 점수가 달라지며 이로 인해 심사결과와 신뢰도가 낮아질 수 있음을 이야기하였다. 또한, 댄스스포츠 심사결과를 분석한 Kim & Jeong(2006), Kim et al.(2012), Kim & Jeon(2018), Premelč et al.(2019), Lee & Kim(2020), Kim & Lee(2021)의 연구에서는 심사 및 순위 결정 과정에서 심판의 국적, 경력과 같은 개별 특성과 심사 위치, 심사해야 하는 대상자의 수 등 심사 조건에 따라 부여하는 점수와 등위 판정에 대한 일치도가 심판 간에 차이가 있음을 확인하였다. 특히, Kim & Jeong(2006), Kim et al.(2012)의 연구에서는 본 연구의 슛컷 사용 의견 분석결과에서 나타난 음수의 일치도와 유사하게 역상관 관계가 일부 종목 일치도 결과에서 나타나 현재 사용 중인 심사 기준과 채점 방법에 문제가 있으며 이를 해결하기 위한 노력이 필요함을 강조하였다.

본 연구와 선행연구에서 나타나는 심판 간 심사의견과 점수의 차이, 이로 인한 낮은 신뢰도 결과는 현재 진행되고 있는 심사제도에 부족함이 있음으로 해석할 수 있다. 이는 심판의 주관에 의존하여 결과를 결정하는 분야에서는 매우 심각한 문제이다. 왜냐하면, 같은 대회 내에서 동일한 심사 방법을 사용하여도 남·녀 경기, 예선과 결선, 나아가 해가 바뀌어 대회가 계속 진행되는 동안 일정 수준 이상의 심사결과 신뢰성을 확보하지 못할 수 있음을 의미하기 때문이다(Suk & Lee, 2022). 일정하지 못한 대회 결과의 공정성과 신뢰성은 단순히 대회 권위 및 경기결과와 가치 하락뿐 아니라, 심판에 대한 선수의 신뢰를 하락시켜 선수가 선보이는 경기력에도 영향을 미칠 수 있는 중요 요인으로 보고 되고 있다(Seo et al., 2022). 이에, 주관적 판정을 사용하는 체육, 무용, 스포츠 분야에서는 이러한 문제를 해결하고, 결과의 신뢰도 확보를 위해 공통적으로 동일 심사 자격을 갖춘 다수의 심판이 심사에 참여하고, 각 분야의 특성에 맞추어 제작된 심사 지표, 도구 등을 사용하고 있다(Ahn & Oh, 2007; Cho & Choi, 2015; Lee et al., 2016; Premelč et al., 2019). 브레이킹 역시 스포츠로서의 체계화 과정에서 이를 반영하여 심사제도 구축에 많은 노력을 기울여 왔다. 가장 먼저 국제기구 산하에 심판 관련 기관과 심판 자격제도를 설립하여 각 국가별 심판 자원을 확보하고, 정기적인 심판 교육을 진행하여 다수의 심판이 심사에 참여할 수 있도록 하고 있다(KFD, 2022a; WDSF, 2021b). 그리고 기술적 요인 외에 개성, 창의성, 음악성 등 다양한 요인에 대한 종합적인 능력이 요구되는 브레이킹 배틀의 심사를 위해 새로운 트리비움 밸류 시스템과 트리폴드라는 새로운 심사 방법을 개발하여 현장에 도입하였다(Fahr et al., 2018; USA Dance, 2022; WDSF, 2019). 이 심사 방법은 브레

이킹에서 나타날 수 있는 여러 가지 요인을 종합하여 대상자의 차이를 명확하게 구별하고자 심사 준거와 그에 따른 점수 척도를 포함하도록 제작되어 있다. 하지만 심사결과와 신뢰성 확보를 위해 기술인 이러한 노력에도 불구하고 본 연구결과로 미루어 볼 때 트리비움 밸류 시스템과 이를 사용한 현재 심사제도는 브레이킹 배틀을 구성하는 다양한 요인에 의한 차이를 명확하게 심사하는데 적절하지 못한 것으로 보여지므로 심사결과와 신뢰성 확보를 위해 보다 폭 넓은 연구적 접근과 이를 기반으로 한 제도적 개편이 필요하다.

본 연구의 대상인 브레이킹은 하나의 문화적 요소로 출발한 만큼 스포츠로서 많은 체계화를 거쳤음에도 여전히 높은 자유도를 가지고 있다. 현재 배틀에 사용하는 음악은 별도의 규정 없이 주최 측에서 자유롭게 선택하도록 하고 있으며(Fahr et al., 2018, WDSF, 2021a) 이로 인해, 배틀의 매 라운드에 다른 속도(Beats Per Minute : BPM) 종류의 음악이 제공되어 선수들은 음악에 맞추어 적절한 기술을 즉흥적으로 사용해야 한다. 브레이킹 기술의 경우 '탑 락(Top Rock : 서 있는 상태에서의 동작)', '다운 락(Down Rock : 신체 일부가 바닥에 닿은 상태에서 실시되는 동작)', '파워 무브(Power Move : 원시력을 이용한 회전 동작)', '프리즈(Freeze : 전체적인 움직임 안에서 멈추는 동작)'의 4가지 요소로 구분(KFD, 2022b)하고 있으나 선수별로 신체적 특성과 숙련도, 선호도에 따라 다양한 기술 조합이 존재한다(Cha, 2021; Moon & Kim, 2021; Park et al., 2021; Wei et al., 2022; Yang et al., 2022). 더욱이, 브레이킹에서는 같은 동작과 기술을 반복적으로 사용하는 것을 감점요소로 반영(Fahr et al., 2018)하기 때문에 배틀의 매 라운드와 선수별로 기술과 춤은 매우 여러 가지 형태로 나타날 수 밖에 없다. 이는 음악에 따라 수행되는 대상자의 기술과 춤이 가지고 있는 다양성과 자유도가 높음을 의미하며 이를 주관적으로 심사하는 심판의 차이와 상호작용을 통해 심사결과에 영향을 미칠 수 있음을 시사한다. 따라서, 이러한 상황에서 사용하게 되는 심사 기준은 심판 간 동의의 통해 대상자의 수행능력에서 나타나는 차이만을 명확하게 구분할 수 있어야 한다(Frey et al., 2012; Pinheiro, 1994). 하지만, 본 연구의 결과에서 나타난 낮은 심판 간 신뢰도와 일치도는 현재 브레이킹 심사를 위해 개발된 트리비움 밸류 시스템에 부족함이 있음을 의미하며, 보다 나은 심사를 위한 개선 및 보완이 필요할 것으로 생각된다.

이를 위해, 먼저 심사 항목의 구체화 과정과 그에 따른 점수 체계를 구축할 필요성이 있다. 현재 체조, 피겨스케이팅과 같은 종목이나 프리스타일 스키(Freestyle ski), 스노보드 프리스타일(Snowboard Freestyle)과 같은 종목에서는 경기상황에서 나타날 수 있는 기술들을 분류하고, 기술의 난이도와 특성에 따라 점수체계를 세분화하여 대회 심사에 적용하고 있다(FIG, 2022; FIS, 2022a, 2022b; ISU, 2022a). 이는 단순히 1회성으로 제작되어 사용하는 것이 아닌 종목의 자유도와 기술의 발전에 따라 발생하는 새로운 기술들을 지속적으로 추가하고, 기술에 따른 점수체계를 수정·갱신해 현장 심사에서 사용하고 있다. 브레이킹 역시 적극적으로 기술을 체계화하고 그에 따른 점수체계를 보다 세분화하는 작업이 필요할 것으로 보여진다. 이러한 과정을 통해 개발된 심사 도구를 현장에 심사에 적용함에 있어 사전에 심판 교육과 대회 준비과정을 통해 심판 간 정기적으로 의견을 조율하고, 심사 기준을 보다 명확하게 일치시키는 과정이 함께 이루어진다면 본 연구결과에서 나타난 심판 간 심사의견과 결과의 차이를 줄일 수 있을 것으로 사료 된다.

또한, 현재 단순히 모든 항목을 5명의 심판이 총체적(Holistic)으로 심사하는 방법을 보완하여 심사 준거에 따라 각각의 항목별로 개별 심사위원을 배정하고, 그에 따른 분석적(Analytic) 심사가 함께 이루어지게 하는 것이 보다 신뢰성 있는 심사결과를 얻을 수 있는 방안이 될 것으로 생각되며 이에 따른 최적 심판 수에 대한 연구가 필요하다. 브레이킹 배틀 심사 상황에서 심판은 40초~1분의 짧은 시간 내에 배틀을 펼치고 있는 두 선수에 대해 각각 트리비움 밸류 시스템에 배정되어있는 6가지 항목에 대한 모든 심사를 하여야 하며, 이때 숫자 버튼을 이용한 감점과 가점에 대한 점수까지 부여하여야 한다. 이로 인해 매우 짧은 시간 내에 많은 항목에 대한 심사를 진행해야 하기에 심판에게 부담을 주게 되고, 이는 제대로 된 심사를 결과를 얻는데 악영향을 미칠 수가 있다. 현재 체조, 피겨스케이팅과 같은 종목에서는 심사 항목을 세분화하고 각 영역에 맞는 전문성을 갖춘 다수의 심판이 심사에 참여하도록 하고 있으며, 이를 위해 9심~12심제 등의 다양한 심사 조건을 종목과 심사 항목의 특성에 맞추어 선택하여 사용하고 있다(FIG, 2020; ISU, 2022b). 이에 비해 브레이킹 배틀 심사에 활용하고 있는 5심제는 심사 도구의 복잡성에 비교하였을 때 올바른 심사를 진행하는데 무리가 있는 것으로 보여진다. 따라서 브레이킹을 위해 개발된 심사 도구의 수정 및 보완작업과 함께 이를 사용할 경우 필요한 최적의 심판 수를 찾는 연구가 이루어져야 할 것으로 생각된다.

브레이킹은 비교적 최근에 스포츠로의 체계화 과정을 거치고 있는 신생 종목으로 타 종목들에 비교했을 때 상대적으로 그 역사가 짧아 현재 새롭게 제도적 구축을 위한 노력이 지속적으로 기울여지고 있다. 이를 위해 국제기관이 설립되고, 각 국가별 협회들이 신설되어 브레이킹을 위한 각종 규정이 만들어지고 있으며, 올바른 대회 운영과 객관성 있는 심사를 위한 심사제도 구축 및 정기적 심판 교육과정이 함께 이루어지고 있다(Fahr et al., 2018; KFD, 2022a; WDSF, 2021b, 2022b). 하지만 이러한 노력에도 불구하고, 첫 번째 메가 스포츠 이벤트 정식종목으로 이루어진 2018년 부에노스 아이레스 유스 올림픽의 브레이킹 심사결과는 매우 낮은 심판 간 신뢰도 및 일치도를 보여 새롭게 도입된 브레이킹 심사제도에 보완 필요성이 있음을 본 연구를 통해 확인할 수 있었다. 2024년 파리 올림픽을 준비하는 시점에서 브레이킹이 하나의 문화적 요소가 아닌 스포츠 종목으로서 인정을 받기 위해서는 이러한 문제를 해결해야 할 것이다. 따라서, 연구와 현장 전문가들이 소통과 협업하여 브레이킹 배틀 심사를 위해 개발된 심사 도구와 이를 사용하는 심사 조건에 대한 학문적 검증과 더불어 현장 결과를 분석한 실증적 연구를 통해 현재 심사제도의 문제점을 찾고 보완하는 노력이 지속되어야 할 것으로 생각된다.

결론 및 제언

본 연구는 브레이킹 배틀의 심사를 위해 개발되어 현장에서 활용 중인 트리비움 밸류 시스템의 특성을 소개하고, 이를 사용하여 진행된 2018년 부에노스 아이레스 유스 올림픽의 심사결과를 분석하여 심판 간 신뢰도와 일치도를 확인하는 목적으로 수행되었다. 이를 위해 2018년 부에노스 아이레스 유스 올림픽의 브레이킹 B Boy, B Girl 개인전의 예선, 결선 결과를 분석하였으며 연구결과를 통해 얻은 결론은 다음과 같다.

연구결과 B Boy, B Girl의 예선과 결선 모두 심판 간 신뢰도는 낮았

으며 심사 항목과 대상별로 신뢰도 경향에 차이가 있는 것으로 나타났다. 또한, 배틀의 승·패 여부를 결정하는 심판 간 의사결정 결과와 세부 항목별 심사의견 및 감점, 가점을 위한 심판 간 의견에서 모두 낮은 일치도를 보였다. 이러한 연구결과는 현재 진행되고 있는 트리비움 벨류 시스템을 이용한 브레이킹 배틀 심사에서 심판 간 의견과 점수에 많은 차이가 있음을 의미한다. 본 연구에서 나타난 낮은 심판 간 신뢰도와 심사의견의 일치도는 새롭게 메가 스포츠 이벤트 정식종목으로 인정받은 브레이킹이 스포츠의 한 분야로서 발전하는데 저해 요소로 작용할 수 있다. 따라서, 보다 나은 심사결과를 얻을 수 있도록 현장과 연구의 협업을 통한 다각도의 노력이 이루어져야 할 것이다.

이를 위한 후속연구를 위한 제언은 다음과 같다. 먼저 현재 브레이킹 배틀을 위해 개발되어 사용되고 있는 트리비움 벨류 시스템을 보완하여 기술 및 심사항목을 체계화하고 그에 따른 점수체계를 보다 세분화하는 작업 선행되어야 할 것이다. 또한, 트리비움 벨류 시스템을 사용한 심사를 진행할 때 심판은 승자와 패자의 선정과 두 선수에 대한 항목별 점수 배점, 감점과 가점 요소에 대한 심사까지 짧은 시간 동안 지나치게 많은 요인에 대한 심사를 진행해야 하는 것으로 판단된다. 이에 브레이킹을 위해 개발된 심사도구를 사용할 경우 결과의 신뢰성을 확보할 수 있는 최적의 심판 수를 찾는 연구가 필요할 것이다. 이러한 과정을 통해 지속적으로 현장에 심사제도의 보완과 새로운 대안을 제시하는 연구가 진행된다면, 보다 나은 심사결과를 얻을 수 있는 심사제도 구축 및 운영에 도움을 줄 수 있을 것이다.

CONFLICT OF INTEREST

논문 작성에 있어서 어떠한 조직으로부터 재정을 포함한 일체의 지원을 받지 않았으며 논문에 영향을 미칠 수 있는 어떠한 관계도 없음을 밝힌다.

AUTHOR CONTRIBUTION

Conceptualization: JS Suk; Data curation: JS Suk & KB Jun; Formal analysis: JS Suk; Methodology: JS Suk; Project administration: JS Suk & KB Jun; Visualization: JS Suk; Writing-original draft: JS Suk; Writing-review&editing: JS Suk & KB Jun

참고문헌

- Ahn, K.-S., & Oh, Y.-J. (2007).** Inter-judges reliability of Korean dance performance assessment. *The Korean Journal of Dance*, 52, 243-254.
- And8.dance. (2022a).** Battle Reports. Retrieved from <https://and8.dance/en/stats>
- And8.dance. (2022b).** And8 judge, every decision counts. Retrieved from <https://and8.dance/en/tools/judge/voting>
- Anshell, M. H. (1990).** An information processing approach to teaching motor skills. *Journal of Physical Education, Recreation & Dance*, 61(5), 70-75.
- Cha, Y.-M. (2021).** Current status of breaking dance training and future challenge. *Research of Dance Education*, 32(3), 125-143.
- 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, J. H. (1999).** A review of extraneous variables on subjective sports skill tests. *Korean Journal of Sport Pedagogy*, 6(1), 105-119.
- Choi, Y. S., & Kang, S.-J. (2004).** The effect of trimmed mean method on the fairness of dance performance assessment. *The Korean Journal of Physical Education*, 43(3), 913-924.
- Choi, Y.-S. (2010).** Performance assessment for modern dance competitors of Dong-A concours (2002~2009). *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 12(2), 23-33.
- Choi, Y.-S. (2011).** Analyzing the judges' rating in modern dance competitors of Dong-A concours through many-faceted Rasch model. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 13(2), 53-62.
- Fahr, D., Gopie, K., Graham, S., Grigo, P., Ma, G., Robitzky, N., ... & Shaughnessy, T. (2018).** *Buenos Aires 2018 Youth Olympic Games WDSF rules and regulations manual: Breaking*. Lausanne, Swiss: WDSF.
- FIG(Fédération Internationale de Gymnastique). (2020).** *Gymnastics for all manual*. Lausanne, Swiss: Author.
- FIG(Fédération Internationale de Gymnastique). (2022).** *2022-2024 code of points*. Lausanne, Swiss: Author.
- FIS(Fédération Internationale de Ski). (2022a).** FIS tricks list: Freestyle. Retrieved from <https://www.fis-ski.com/DB/freestyle-freeski/freeski/trick-list.html>
- FIS(Fédération Internationale de Ski). (2022b).** FIS tricks list: Snowboard. Retrieved from <https://www.fis-ski.com/DB/snowboard/park-and-pipe/tricks-list.html>
- Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012).** Defining authentic classroom assessment. *Practical Assessment, Research, and Evaluation*, 17, 2.
- ISU(International Skating Union). (2022a).** *Judging system: Technical panel handbook*. Lausanne, Swiss: Author.
- ISU(International Skating Union). (2022b).** ISU judging system. Retrieved from <https://www.isu.org/figure-skating/rules/fsk-judging-system>
- Jun, K. B., & Suk, J. S. (2022).** Interrater reliability and agreement analysis according to the characteristics of the breaking battle evaluation. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 24(3), 97-114.
- Kang, S.-J. (2011).** How to estimate inter-rater reliability. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 13(1), 1-8.
- KFD(Korean Federation of Dancesport). (2022a).** *Breaking referee rules*. Seoul: Author.
- KFD(Korean Federation of Dancesport). (2022b).** Introduction: Breaking. Retrieved from http://www.kfd.or.kr/bbs/content.php?co_id=DanceSports_in2
- Kim, E.-J., & Jeong, J.-O. (2006).** Judges' objectivity of DanceSport competition using the skating system. *The Korean Journal of Physical Education*, 45(3), 469-479.
- Kim, H.-D., Kwon, S.-R., & Chea, J.-S. (2012).** Verification of rating agreement of grading judgement for domestic and International Dance Sport Tournament Judges. *The Korea Journal of Sports Science*, 21(3), 1319-1329.
- 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., & Lee, S. (2021).** Analysis of inter-rater agreement of judges in Latin American dance sport. *The Korean Journal of Physical Education*, 60(5), 419-426.
- Landis, J. R., & Koch, G. G. (1977).** The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lee, S.-H., & Kim, M.-S. (2020).** Analysis of inter-rater agreement of latin american and modern dance sport. *Korean Journal of Sport Science*, 31(4), 830-839.
- Lee, T.-K., Kwak, J.-H., Yang, H.-W., & Lee, H.-J. (2017).** Analysis of reliability and error sources of judges' rating in rhythmic gymnastics. *Korean Journal of Sport Science*, 28(2), 337-350.
- Lee, T.-K., Lee, H.-J., Chung, J.-E., & Yang, H.-W. (2016).** Analysis of judges' judging in figure skating. *Korean Journal of Sport Science*, 27(4), 756-769.
- Lee, W.-J. (2022).** The social significance of breakdance battle. *Research of Dance Education*, 33(1), 45-56.
- Moon, B.-S., & Kim, H.-N. (2021).** A study on the preparation plan for the Paris Olympics' breaking dance athletes. *Research of Dance Education*, 32(4), 133-149.
- Oh, S., & 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.
- Oh, S., Kim, S., & Lee, Y. (2001).** An objectivity of a performance assessment instrument in dance. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 3(1), 81-87.

- Park, H., Yoo, H., & Lim, D. (2021).** The controversy over breakdancing becoming an Olympic sport. *Journal of Education & Culture, 27*(2), 509-535.
- Pinheiro, V. (1994).** Diagnosing motor skills—A practical approach. *Journal of Physical Education, Recreation & Dance, 65*(2), 49-54.
- Premelč, J., Vučković, G., James, N., & Leskošek, B. (2019).** Reliability of judging in DanceSport. *Frontiers in Psychology, 10*, 1001.
- Seo, E. K., Kim, D. H., & Lee, S. K. (2022).** The effect of gymnastics athletes recognition of referee judgment on psychology and competition. *Korean Society for the Study of Physical Education, 27*(2), 137-150.
- Suk, J. S., & Lee, J. H. (2022).** Ski level 2 objectivity trend due to changes in evaluation system. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science, 24*(2), 17-38.
- USA Dance. (2022).** *USA Dance breaking for gold USA: Breaking Dancesport rulebook*. Mount Royal, NJ: Author.
- WDSF(World DanceSport Federation). (2017.9.23.).** Trivium value system. Retrieved from https://www.worlddancesport.org/News/WDSF/Trivium_Value_System-2486
- WDSF(World DanceSport Federation). (2019).** *WDSF rules and regulations manual: Breaking* (Version 1). Lausanne, Swiss: Author.
- WDSF(World DanceSport Federation). (2021a).** *WDSF competition rules*. Lausanne, Swiss: Author.
- WDSF(World DanceSport Federation). (2021b).** *WDSF operating policy for the attribution of judge's licenses*. Lausanne, Swiss: Author.
- WDSF(World DanceSport Federation). (2022a).** WDSF rules & regulations. Retrieved from <https://www.worlddancesport.org/rule/competition/general>
- WDSF(World DanceSport Federation). (2022b).** Judging systems. Retrieved from https://www.worlddancesport.org/Rule/Competition/General/Judging_Systems
- Wei, M. T., Yang, Z., Bai, Y. J., Yu, N., Wang, C. X., Wang, N., & Cui, Y. S. (2022).** Shaping future directions for breakdance teaching. *Frontiers in Psychology, 13*, 952124.
- Yang, Z., Bai, Y. J., & Wei, M. T. (2022).** The importance of creativity in the sportification of breakdance. *Frontiers in Education, 7*, 855724.

2018 부에노스 아이레스 유스올림픽 브레이킹 배틀 심판 간 신뢰도 및 일치도 분석

석진수¹, 전경배²

¹용인대학교, 강사

²한양대학교, 박사과정

[목적] 본 연구는 2018년 부에노스 아이레스 유스올림픽부터 새롭게 정식 종목이 된 브레이킹 심사결과에서 나타나는 심사 경향과 심판 간 신뢰도 및 일치도를 분석하여 보다 나은 브레이킹 심사를 위한 기초자료를 마련하는데 있다.

[방법] 2018년 부에노스 아이레스 유스올림픽 브레이킹 B Boy, B Girl 개인전의 예선과 결선 결과를 연구자료로 수집하였으며, 수집된 자료는 Window Excel 2021을 사용하여 통계 방법에 맞게 정리하였다. 항목별 심사점수의 특성과 심판 간 신뢰도 분석을 위하여 기술통계 및 유목내상관계수를 산출하고, 승자·패자 선정과 항목별 심사의 심판 간 일치도는 플레이스 카파를 산출하여 확인하였으며 모든 자료처리에는 Windows SPSS 26.0를 사용하였다.

[결과] 연구 결과 B Boy, B Girl 모두 예선·결선에서 심판이 부여한 점수에 심판 간 차이가 있었으며, 심판 개인 점수에서 편차가 큰 것으로 분석되어 모든 항목에서 매우 낮은 심판 간 신뢰도를 보이는 것으로 나타났다. 또한, 승자·패자 선정 결과와 항목별로 심판 간 부여한 감점 및 가점 요소에 대한 심판 간 의견의 일치도 역시 매우 낮은 것으로 분석되었다.

[결론] 연구 결과를 종합하여 살펴보았을 때 동일한 심사기준이 반영된 트리비움 밸류 시스템을 가지고 브레이킹 심사에 참여하는 심판 간 심사의견에 많은 차이가 있는 것으로 나타나 이러한 차이를 줄일 수 있는 개선 노력이 필요할 것으로 보여진다. 또한, 현재 브레이킹 심사에 사용하고 있는 트리비움 밸류 시스템은 심사과정에서 고려해야 하는 사항이 많고 채점체계가 복잡하여 현재의 심판 수는 적절하지 않은 것으로 보여진다. 따라서 보다 나은 심사를 위해 심사제도의 개편과 최적의 심판 수를 확인할 필요성이 있을 것으로 사료 된다.

주요어

브레이킹 배틀, 트리비움 밸류 시스템, 심판 간 신뢰도, 심판 간 일치도