Global Hip Hop Studies
Volume 4 Number 2
 https://doi.org/10.1386/ghhs_00095_1
Received 1 June 2022; Accepted 6 February 2024; Published Online July 2024

YONATAN ASHER VEXLER
Child of this Culture Foundation

CANDY FOELIX
Child of this Culture Foundation

CINDY FOLEY
Child of this Culture Foundation

Breaking for gold: Who can judge me?

ABSTRACT

Despite the central importance of judging in breaking competition, little quantitative research on how different people judge and score breaking battles has yet been done. This article explores variables that potentially influence performance evaluation. Statistical analyses of survey data show how these variables can affect battle results. Findings concerning adjudicator certification application vetting and course teaching recommend to not discriminate by applicant veterancy yet focus on thorough curricula, customized per country. More judge certification course time should also focus on the proper comprehension of performances' interpretative and artistic qualities. The final exam should be what defines who can be a good judge. Implications reach as far as education for the Olympic competition audience. Such curricula can use videos as described herein. Suggestions for further study are also discussed.

KEYWORDS

dance
battle
scoring
adjudication
Olympics
certification
education
trivium

Promoted as an Olympic sport, breaking could become increasingly popular across the globe. With the potential for an increase in competition battles and tournaments, new demand would require more judges to deliberate

their winners. To certify Olympic breaking competition judges, the World DanceSport Federation (WDSF) has implemented a short workshop and an online exam for prospective judges. The certification process begins with a questionnaire about one's breaking dance and judging veterancy and ends with an exam, which includes questions about breaking history and dance foundation, as well as questions about the judging models and system interfaces, to ensure that certified judges all have similar understandings of the dance and its adjudication.

Judgement of dancesport events and, specifically, breaking battles has been emphasized as of the utmost importance (Anderlucci et al. 2021; Li and Vexler 2019; Fogarty 2019). Judges have large and lasting impacts on breaking and its evolution. Judges' feedback can bring about rapid improvement in dancer performance and even set new trends, such as when judges reward competitors 'who take more risks' or reward teams who 'have much cleaner endings' (Fogarty 2019: 424). Just as the judge certification itself is an educational activity, the role of the breaking judge goes beyond battle scoring. This is a story of adjudication and education, with an aim to explore how similarly different cohorts of battle viewers comprehend, interpret and score this competitive dance. Understanding such similarities and differences can provide insights for judge application vetting and course teaching.

If only veteran breakers should be accepted for certification, how many years of experience are considered necessary to develop a thorough understanding of the dance's culture and competition? Certification course teachers should also be aware of which qualities are harder to judge, so that more explanation can be provided on those details. Veterancy and comprehension may also differ by country. Young amateurs in one region might have already achieved enough of an understanding to pass the certification exam, while older professionals from other regions may grow to have vastly different interpretations on particular aspects of breaking performance. WDSF breaking now requires rapid growth of the worldwide certified judge pool, yet proficiency and excellence are also a sport priority, necessitating this new chapter on adjudication and education. We start by trying to map the complex relationships between judging experience (years spent judging competitions), dance veterancy (years spent breaking) and geographical location on judging decisions, because it is expected that people from different countries, as they become more experienced dancers and judges, eventually reach similar understandings of how to interpret the battles they see.

LITERATURE REVIEW

Despite adjudication being a major concern for both dancesport and breaking competition (Anderlucci et al. 2021; Li and Vexler 2019; Fogarty 2019), it has rarely been studied. Searches in Google Scholar, Web of Science, ProQuest and other databases, using combinations of the keywords 'breaking', 'dance', 'breakdance', 'competition', 'judge' and 'judging' found very few relevant articles. Most were qualitative descriptions of breaking competition, only briefly mentioning some aspects of judging. When sharing breaking competition history and culture, sometimes of a specific region or event, mention of adjudication within articles can be as slight as one sentence. For example, Dalecki's study on the B-Boy City event only shared one datum on judging – 'physical contact between opposing crews is not permitted, and many battle judges

[…] will disqualify teams for initiating contact, though not all judges strictly enforce the tenet' (2011: 366).

Only one English article focuses specifically on issues of breaking competition judging. Fogarty's (2019) account of the realities, complexities and implications of battle judging presents its history and shares exemplary cases of judges' work that highlights their role within events and within the greater breaking community, challenges judges face, the power they wield, and also how breaking events and judging work have changed over time. The next most relevant English article mostly focuses, more specifically, on education. Yang et al.'s (2022) article on the importance of 'creativity' stresses how necessary it is for teachers, students and judges to have a clear and common understanding of this essential competition scoring criterion; however, they did not provide data on specific discrepancies or prevalence. Yang et al. (2022) frequently cited Shimizu and Okada (2018), who studied breakers' creativity practice processes. There is a growing body of research on breaking, but quantitative studies on breaking competition are still scarce, and studies focused on judging are even more rare. A doctoral candidate at the University of Michigan has been collecting battle judging data and will hopefully soon make a great contribution to this budding field.

Research on judging is similarly scarce in Asia. For example, a search on the CNKI database using combinations of the keywords '街舞', 'breaking', '赛事', '裁判', '评分' and '竞技' returned many articles relevant to breaking competition, practice and education, but only two articles were found to substantially discuss adjudication and they were mainly just introducing the Olympic judging system to their Chinese readership.

Vexler (2021) summarized the literature on breaking and did not report judging as any significant portion of it. This summary did, however, note the challenges of searching for studies on breaking. It is hard to sort articles on hip hop dance or street dance that may have studied breaking. For example, Sato and Hopper's (2021) article on the reliability of judges evaluating a hip hop dance movement mentions 'break dance' and the Olympic event judging model within its introduction and discussion sections, but then refers to 'hip-hop dancers' and hip hop competition judges within the subsection on study participants. The move used for their study is not one specific to breaking and it is unclear whether the reported reliability regards breaking judges.

The literature would benefit from a distinction between terms because referring to them as synonyms reduces their usefulness (Baker 2012). Scholars can help readers identify studies about breaking by only using this distinct word. We therefore write this article about breaking, breakers, and breaking judges, without confusing these with hip hop dance or hip hop dancers.

Only one past study did report on the reliability of breaking judges, specifically and unambiguously. Vexler's (2021) thesis included the scoring of dance videos by professional breaking competition judges and statistical evidence was presented to support the claim of judge reliability. The OUR system of judging was used, with its five categories: 'foundation' (dancing with musicality, flavour and formula); 'originality' (performing unique characters, styles, moves or creative concepts); 'dynamics' (showing strength, speed, balance, explosive energy, flexibility and illusions of risk); 'execution' (confident, clean, non-repeating motion); and, 'battle' (breaking opponents' confidence by responding to their dance with something better). Slight discrepancies were

found in how the judges scored specific categories, but since the average total scores were so similar and statistical tests found that variance in dancer performance scores were not due to the judges' differences, it was summarized that 'all five judges share a similar value system, as well as carry similar knowledge and experience' (Vexler 2021: 70). Noticing this, one starts to ponder about cases in which judges' scores are significantly dissimilar, how various dancers and audiences interpret the battles that they see, and which cohorts – when using the Olympic judging system – are statistically reliable.

No studies have yet surveyed how different people understand the scoring of competitive breaking using the Trivium system. The Trivium system categorizes artistic qualities (encompassing creativity and personality criteria), physical qualities (encompassing technique and variety criteria) and interpretative qualities (encompassing performativity and musicality criteria). For each of the six criteria, 'judges have an allocated fader on their hand-held Trivium device' (Fahr et al. 2018: 13). The faders are on a scale of four white lines from the start line, implying five main points of measure (with their fractions in between, like a ruler). When dancers demonstrate a more diverse range of movement vocabulary with minimal repetition the variety fader can be moved further from its starting line to give that dancer a comparatively higher variety score.

With Olympic breaking still so new in the world, even experienced breaking competition judges might not yet see battles similarly through the lens of the Trivium system. Fahr et al.'s (2018) manual describes how this system values 'technique' by breakers' differing degrees of athleticism, dexterity, flexibility and spatial control. 'Variety' is evaluated by noticing movement vocabulary dimension and approach. 'Performativity' includes values of composition, narrative and emotion. 'Musicality' is valued by how coherently motions are performed in syncopation, phrased by the music and accentuate its various sounds. 'Creativity' is valued by noticing moves that advance pre-existing ideas and 'personality' includes values of stage presence, charisma and character. While some viewers might easily identify the same values of a dance in a battle round, other judges might disagree, perhaps due to dissimilar understandings of terms and how they appear in dance. Some evaluation criteria could be similarly understood by everyone while other criteria receive wildly varying scores due to how differently people understand them. The process of getting everyone on the same page is a process of education. It is important to know at what stage of this education our dancers and judges are on, to then provide suggestions for facilitating the next steps.

Studies with substantial information on judging have always connected issues of performance evaluation to dance education (see Fogarty 2019; Vexler 2021; Yang et al. 2022). To contribute a fresh angle on this trend and explore areas of investigation that have not yet been researched, this study links dancer experience, judging experience and country location with judging decisions and is guided by the following questions:

1. How do viewers' differing years of breaking and judging experience vary the scores that they give to breaking preselection solo performances and battle rounds?
2. How does the country in which one develops as a dancer predict judging decisions?
3. Which evaluation criteria have the most score variance?

PILOT STUDIES

A pilot study, conducted in 2020, tested the hypothesis that veteran breakers share enough of the same knowledge to be reliable judges. 212 respondents, grouped by years of dance experience, were asked to watch one video of two contestants in battle and decide the winner. All respondents with nine or more years of experience ($n = 65$) chose the same winner. ANOVA tests, used to analyse how similar or different groups of values are, found no significant difference between the scoring of these respondents and the scoring of world-championship breaking judges, $F(1, 63) = 0.014$, $p > 0.1$, and there was likewise no significant variance by country. This implies that those with nine years of experience can be considered veterans and that the judging of local veterans can be as valid as the judging of world-class veterans. Adult respondents (age > 21) with more than three years of breaking experience were also scoring similarly to the veterans ($p > 0.1$), implying that adult breakers might need less dance experience to understand who won a battle.

This online survey was later replicated with a different video and using the Trivium rubric. All respondents with nine years or more dance experience ($n = 65$) chose the same winner again, and again there was no significant difference between the scoring of local veterans and world-class veterans, $F(1, 63) = 0.011$, $p > 0.1$. Veteran members of the breaking community, despite hailing from different countries, seem to be think alike when asked to judge these battles.

Both pilot studies, however, were limited by several factors. The surveys were in English, only deployed via social media to breakers in three countries (United States, Israel and China). Only five respondents had world-championship judging experience, and only one video was scored each time. Also, respondents were grouped by dance experience and age intervals of three years, so no one specific age or any specific years of experience were revealed as the cut-off for veterancy. It could also be that the battles were one-dimensional or too obviously imbalanced, so it was easy enough for dancers with at least some experience to properly evaluate. Testing multivariate regression models using more varied data on more survey items could provide further insights. The research reported here remedies the deficiencies of both pilot studies.

METHODOLOGY

Paradigm

Quantitative study avails objective empirical investigation and tests hypotheses without bias (Bryman 2012; Edmonds and Kennedy 2016), while statistical surveys provide inferences about sample populations (Fowler 2013). This research used a survey that includes multiple battle videos because survey results are more reliable when many items measure the same construct (Shaughnessy et al. 2000; Gliem and Gliem 2003). It was openly distributed to all netizens, available in several languages, because survey results are also more reliable when more and varied populations provide their responses.

Data collection

The original survey (S1) was made in English, on the Typeform platform. It was then translated to Chinese and Hebrew, also hosted on several easy-to-access local platforms. Consisting of 41 items, S1 allowed completion in

parts. Respondents could fill a section, leave it and return later to continue. The first section collected demographic data. The second section was about scoring videos on a 1–5 scale. Respondents were shown three videos of solo dance performances and three videos of single battle rounds. These videos were from small and past events in East Asia that most respondents would not be familiar with (or remember). Section three summarized the main qualities of the Trivium system, provided a brief explanation of each, and respondents scored the videos again, by quality.

Dancers A, B and C performed approximately one minute of preselection solo dance. Dancer D's 65 seconds of performance was one battle round against Dancer E. Dancer E's performance was the response to Dancer D. Dancers F and G, and Dancers H and I, similarly, are battling each other. There are three solo videos and three 1-v-1 battle videos, so the survey included nine total dancers.

A shorter survey was also conducted (S2). Since the Trivium system decides battle winners by side, not by score, and it only matters which dancer won, not by how much, S2 only prompted respondents to choose sides. The solo videos were removed because newly publicized information stated that the Olympic events would not have preliminary rounds, but start straight into battles, seeded similarly to tennis tournaments. As a simulation of Trivium adjudication, S2 is more externally valid. Instead of scoring performances as S1 did, S2 asked which dancers had the better physical, interpretative and artistic qualities in their respective 1-v-1 battles.

Both surveys were shared through social media, websites, newsletters, e-mails and private messages to reach as many respondents as possible. 413 total responses were collected for S1 (266 full questionnaires; 147 questionnaires with unanswered items). A total of 109 full responses were collected for S2, with nine respondents among them having judged world-championship breaking tournaments. Data was not collected for the WDSF breaking levels A and B judging systems because they did not yet exist when this study was conducted.

Data analysis

Cronbach's alpha reliability coefficient should be calculated for Likert-type scale items, especially when a survey is exploratory, as is S1. This test can find how closely related the survey's many items in each portion are to each other. The Trivium scores, for example, should mostly show similar trends of higher scores of each quality for dancers with higher overall average scores, because it would not make sense if someone had extremely high artistic qualities and extremely low physical qualities. Survey results that show a dancer being very creative despite not having the physical ability to move creatively would mean that the survey has a problem. If the items in the scale are not internally consistent, and therefore reliable, further data analysis would be needed. Gliem and Gliem (2003) explained that an alpha of 0.8 or more indicates good reliability.

Multivariate general linear models are used when multiple dependent variables are to be analysed simultaneously (Haase 2011). S1 gathered data on the performance scoring of many videos, also using the Trivium rubric, so the multivariate general linear model can fit all relevant results into one analysis. For S2, binary logistic regression must assess the nominal variable of which side wins the battle (Harrell 2015).

All analyses were done using SPSS 25. The nominal variables of championship judging experience (CJX) and country (CT) were input as fixed factors and the scale variables of age (AG), years of breaking experience (BX) and years of judging experience (JX) were input as covariates for the linear models. Trivium scores of each performance's artistic quality (AQ), physical quality (PQ) and interpretative quality (IQ), the dependent variables of S1, were input as ordinal. In S2, the dependent variable of the winning side was input as either right or left. All analyses used Bonferroni post-hoc pairwise comparisons due to their conservative thresholds (VanderWeele and Mathur 2019).

Ethics

All data are confidential and anonymous. Even the dancers in the videos are referred to only by letter (Dancer A, Dancer B, Dancer C, etc.). Respondents were informed of the research purposes, participated voluntarily and were provided with a contact to email about any concerns. The original videos are openly shared on public platforms, easily accessible and downloadable. After trimming, the videos were embedded in the surveys in such a way that there would be no identifiers, links to the original videos, or links to who originally posted them.

FINDINGS

Reliability

Analysis of S1's personal scoring section (nine items) and Trivium scoring section (27 items) found good internal consistency (Cronbach's alpha = 0.84 and 0.92, respectively). Analysis of all 36 items returned a value of 0.95, so S1's data is reliable and further analyses can continue.

S1 personal scores

386 respondents submitted their personal scoring of the videos on a simple scale of 1–5, with 5 being the highest score. Among these 386 were the scores of eleven respondents who have CJX. Multivariate tests of the nine video scores found statistical significance for most variables (Table 1). JX was significant ($p < 0.5$) for two of the videos, meaning that the average scores of people with more and less JX (including zero JX) were similar for seven performances. BX was significant for six of the videos, meaning that respondents of varying experience all submitted similar average scores for three performances. Scores of five videos were also found to have no significant difference by respondent age (AG), yet all videos' scores varied significantly by CT. No significant difference ($p > 0.5$) was found between the scoring of those who are championship judges and those who are not, for any of the videos. The videos of Dancers H and I had the least variance, with only CT being significant ($p < 0.5$). Contrarily, the video of Dancer F had the most variance, with all effects except CJX being statistically significant. This performance could have been the most difficult to interpret, so scoring was greatly varied even among respondents with CJX. These five predictors were able to explain a large proportion of the variance in personal scoring (all adjusted $R^2 > 0.250$), indicating that the model has strong predicting power.

Most respondents were from the United States ($N = 131$), Israel ($N = 63$) and China ($N = 57$). More variance was usually found among the scores of countries with fewer respondents; however, the personal scoring of Dancers B

Table 1: Personal scoring multivariate tests.

Variable	Test	Value	*F*	Degrees of freedom	*p*
JX	Pillai's Trace	0.322	18.932	9,358	0.000
	Wilks's Lambda	0.678	18.932	9,358	0.000
	Hotelling's Trace	0.476	18.932	9,358	0.000
	Roy's Largest Root	0.476	18.932	9,358	0.000
BX	Pillai's Trace	0.435	30.651	9,358	0.000
	Wilks's Lambda	0.565	30.651	9,358	0.000
	Hotelling's Trace	0.771	30.651	9,358	0.000
	Roy's Largest Root	0.771	30.651	9,358	0.000
AG	Pillai's Trace	0.137	6.298	9,358	0.000
	Wilks's Lambda	0.863	6.298	9,358	0.000
	Hotelling's Trace	0.158	6.298	9,358	0.000
	Roy's Largest Root	0.158	6.298	9,358	0.000
CT	Pillai's Trace	3.136	15.053	117,3294	0.000
	Wilks's Lambda	0.010	19.812	117,2689	0.000
	Hotelling's Trace	7.903	24.063	117,3206	0.000
	Roy's Largest Root	3.339	94.004	13,366	0.000
CJX	Pillai's Trace	0.037	1.534	9,358	0.134
(Yes or No)	Wilks's Lambda	0.963	1.534	9,358	0.134
	Hotelling's Trace	0.039	1.534	9,358	0.134
	Roy's Largest Root	0.039	1.534	9,358	0.134

and F were significantly different ($p < 0.5$) also among these three countries. For Dancers H and I, scores did not vary by BX but did vary by CT.

Looking at dance experience, those with BX > 6 seem more likely to provide similar scores for Dancer E and are also more likely to interpret his performance similarly to how those with CJX do. Although those with BX > 2 are more likely to provide similar scores as championship judges, there is much more variance between and within groups for Dancers F and H. Average scores for Dancer I are most interesting, since most respondents with CJX provided similar scores, but other respondents of similar veterancy had quite different interpretations of her performance (Figure 1). Those with 2 < BX < 15 seem likelier to score Dancer I similarly to those with CJX, but that likelihood is not very high. For Dancer G, only those with BX > 13 were more likely to score like those with CJX. Overall, the average cut-off for personal scoring to likely mimic the personal scoring of those with CJX was BX > 4.

S1 Trivium scores

Multivariate PQ tests ($N = 275$) found all effects significant (Table 2). AG was a significant predictor of score variance for Dancers A, C and E ($p < 0.05$); BX was a significant predictor for Dancers A, C, F, G and H; JX was a significant predictor for Dancers A, C, E, F and G; CT was a significant predictor for

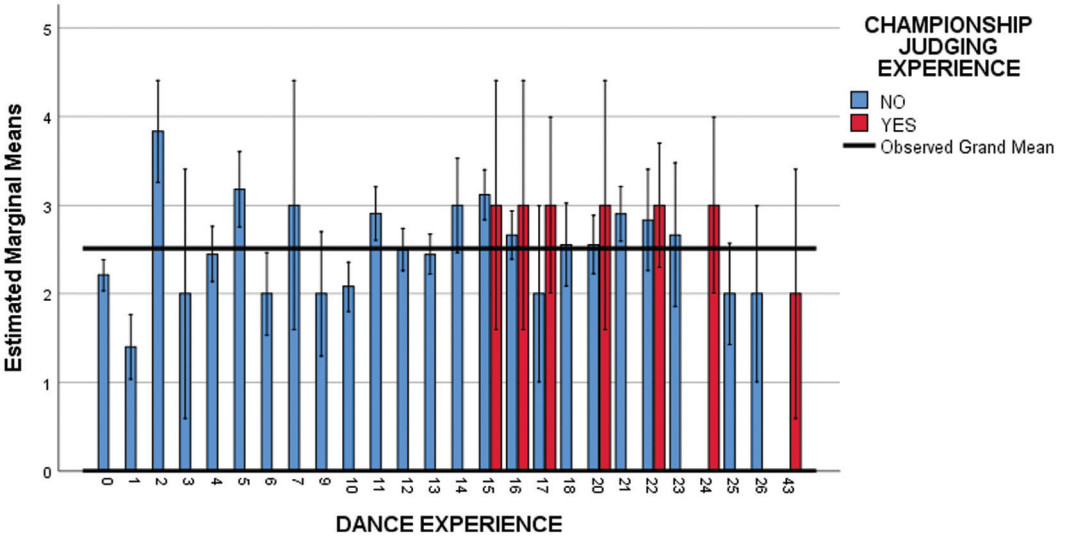

Figure 1: Personal scoring of Dancer I.

Dancers A, B, C, E, H and I; while CJX was significant for Dancers C, D, E, F and G. The PQ scores of Dancer C were most varied, meaning that the quality of this performance is interpreted differently by most viewers. Complementarily, the PQ of Dancers D and I seem easiest to judge, as only one effect on each was found significant. The model explains a large proportion of PQ score variance (all adjusted $R^2 > 0.343$), indicating strong predicting power.

PQ scores for Dancer C vary significantly by all factors. South Koreans scored this quality of her performance poorly, while Americans submitted higher scores, and Americans with CJX submitted even higher scores (Figure 2). A similar situation occurred with the scores of Dancer I. These results indicate that people of various backgrounds and veterancies have different understandings of how to judge the PQ of these dancers' performances. Gender could also perhaps be relevant, since these two dancers appear to be female and people from different countries may have different ideas on how to judge female competitors.

Regarding Dancer D, those without CJX tended to score his PQ higher than those with CJX (Figure 3). The PQ scoring of Dancer F was very varied in groups and between groups, meaning that most respondents disagreed about how well his PQ really was. For Dancer A, those with BX > 7 seem most likely to score like a championship judge. For the performance of Dancer B, the cut-off is at BX > 4. Averaging all cut-off years for the nine dancers results in an estimated BX > 5 required to understand PQ like those with CJX.

Regarding IQ scores (N = 275; 11 with CJX), multivariate tests found all effects significant ($p < 0.05$). The model has strong predictive power (Table 3), considering that social science research often deems R^2 values of 0.20 as able to explain large proportions of variance. AG was significant for Dancers A, B, C, E and H; BX was significant for Dancers A, C, D, F, G, H and I; JX was significant for Dancers A, B, C, D, E, H and I; CT was found significant for all; with CJX significant for Dancers B, F and G.

Table 2: PQ score multivariate tests.

Variable	Test	Value	*F*	Degrees of freedom	*p*
AG	Pillai's Trace	0.162	5.312	9,248	0.000
	Wilks's Lambda	0.838	5.312	9,248	0.000
	Hotelling's Trace	0.193	5.312	9,248	0.000
	Roy's Largest Root	0.193	5.312	9,248	0.000
BX	Pillai's Trace	0.435	21.252	9,248	0.000
	Wilks's Lambda	0.565	21.252	9,248	0.000
	Hotelling's Trace	0.771	21.252	9,248	0.000
	Roy's Largest Root	0.771	21.252	9,248	0.000
JX	Pillai's Trace	0.157	5.148	9,248	0.000
	Wilks's Lambda	0.843	5.148	9, 248	0.000
	Hotelling's Trace	0.187	5.148	9,248	0.000
	Roy's Largest Root	0.187	5.148	9,248	0.000
CJX	Pillai's Trace	0.152	4.952	9,248	0.000
(Yes/No)	Wilks's Lambda	0.848	4.952	9,248	0.000
	Hotelling's Trace	0.180	4.952	9,248	0.000
	Roy's Largest Root	0.180	4.952	9,248	0.000
CT	Pillai's Trace	3.809	15.655	108,2304	0.000
	Wilks's Lambda	0.002	21.663	108,1818	0.000
	Hotelling's Trace	11.275	25.705	108,2216	0.000
	Roy's Largest Root	4.142	88.358	12,256	0.000

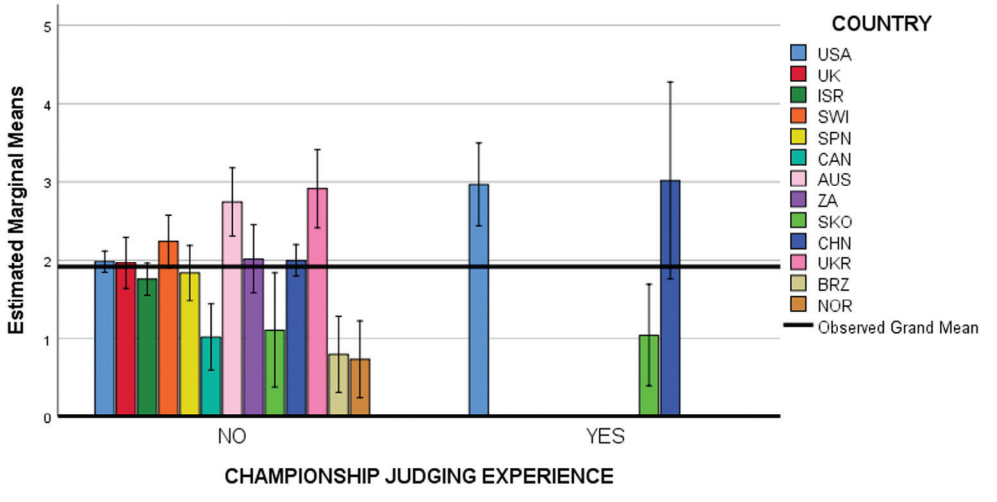

Figure 2: Dancer C's PQ scores.

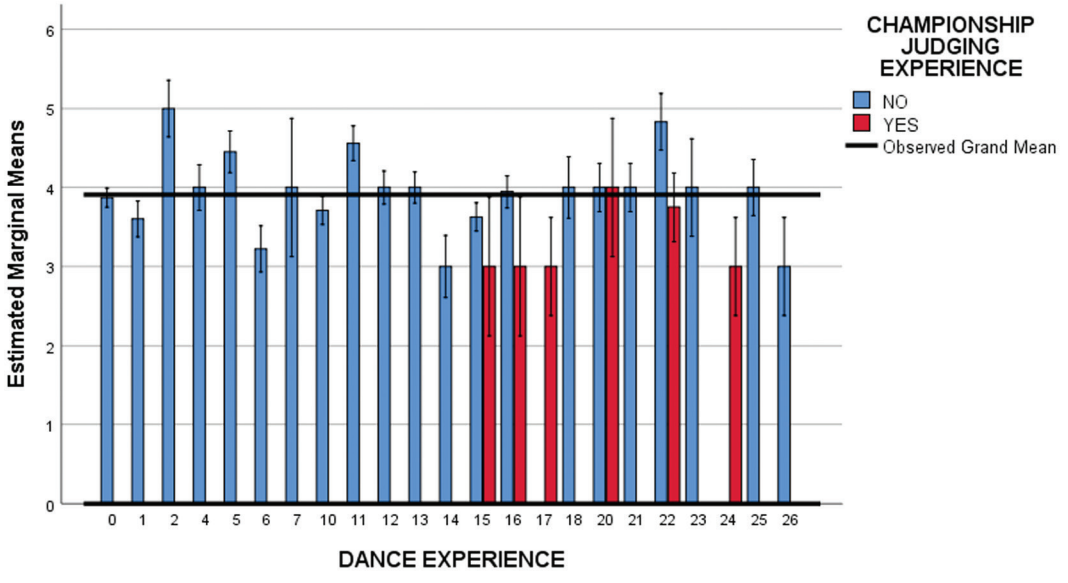

Figure 3: Dancer D's PQ scores.

Table 3: IQ R^2 values.

Dancer	R^2	Adjusted R^2
A	0.561	0.530
B	0.447	0.408
C	0.561	0.530
D	0.358	0.313
E	0.324	0.276
F	0.513	0.479
G	0.401	0.359
H	0.406	0.364
I	0.410	0.368

IQ scores vary by more factors than PQ. Five dancers' IQ have four significant effects and none have less than three, while significant effects on some dancers' PQ were as few as one. This is evidence that IQ is harder to judge than PQ, possibly because most respondents are far from being on the same page regarding what IQ looks like.

Regarding the veterancy cut-off for judging IQ, those with BX > 8 seem more likely to score like a championship judge. For Dancer A, average scores of those with JX < 8 can be very similar to the championship judges', and scores of those with JX > 8 are exactly the same as the championship judges'. For Dancer B, average scores of those with JX = 3 already equal to

those of championship judges with JX = 12 and 14, but the scores of those with JX = 2 are already very close (Figure 4). Calculating all averages, those with JX > 6 are most likely to score IQ like a champion judge. This cut-off is greater than the cut-off for PQ, strengthening the claim that IQ is more difficult to judge.

Regarding AQ scores (N = 275; 11 with CJX), multivariate tests found all predictors significant ($p < 0.05$). The model has strong predictive power (all $R^2 > 0.197$). AQ scores for Dancer A are significantly predicted by AG, BX, JX and CT (all $p < 0.05$); scores of Dancers B, E and F are significantly predicted by JX and CT; scores of Dancers C, D, G are significantly predicted by BX, JX and CT; Dancer H's scores are significantly predicted by BX and CT; while Dancer I's scores are significantly predicted by CT. CJX is significant only for the AQ scores of Dancers D and G ($p < 0.05$). Comparing the number of significant predictors, AQ has fewer than IQ, indicating that less judge characteristics are connected to variances in AQ scores. This could mean that AQ is easier to uniformly judge than IQ.

Regarding the cut-off years for AQ judging veterancy, averages have been calculated to BX > 4 and JX > 3, also strengthening the claim that IQ is the most difficult to judge; however, AQ can also be unusually hard to interpret. For example, only the average scores of those with JX equalling seven, eight and fifteen equal the championship judges' average scores for Dancer G (Figure 5). The AQ scoring of Dancer A was also significantly varied in groups and between groups, meaning that most respondents did not agree about how artistic his performance really was.

Some videos seem harder to score due to varying effects in the model. All three qualities of Dancers A, C, D, E and G significantly vary by many factors. Dancer F's score was more varied when respondents used personal

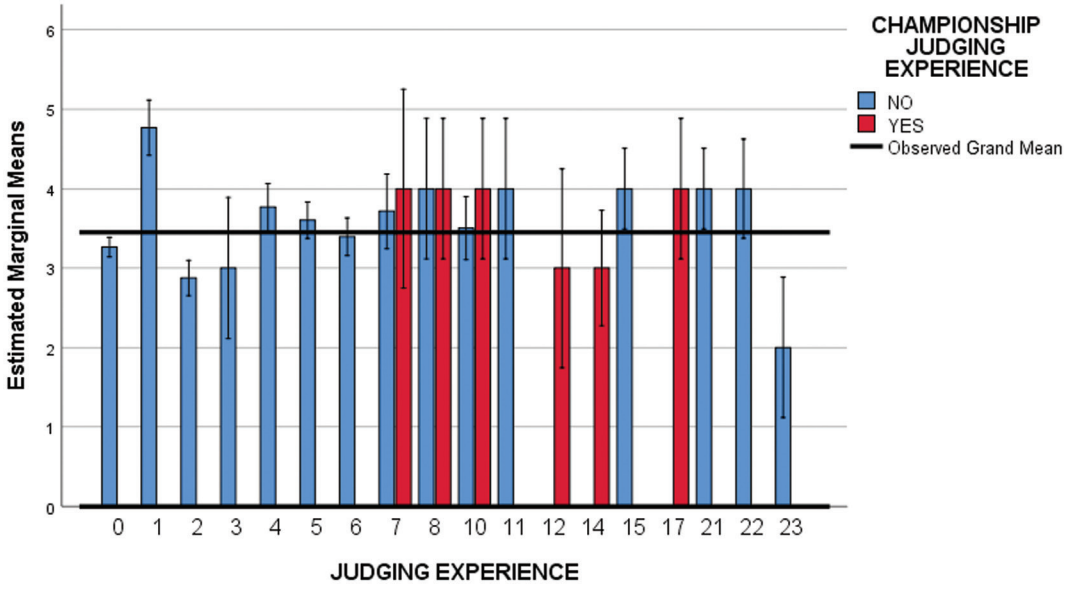

Figure 4: Dancer B's IQ scores.

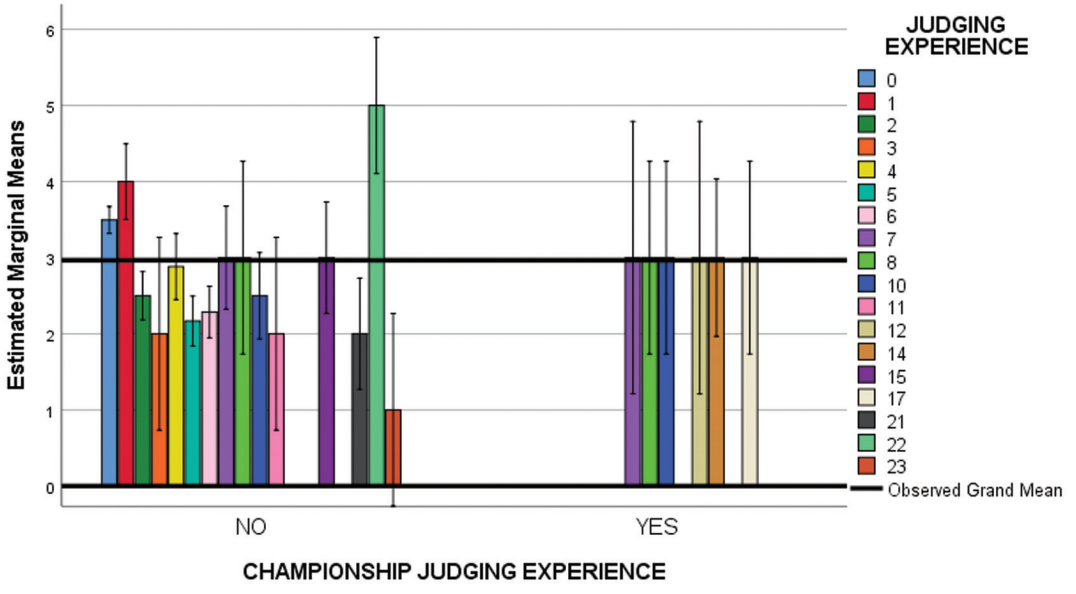

Figure 5: Dancer G's AQ scores.

judgement and became less varied when the Trivium framework was used, but interpreting his performance with Trivium substantially decreased his total average score, because his AQ scores were low. When paying more attention to the specific qualities of the performances, rather than their overall impression, all dancer average scores decreased, but not all became less varied.

Since PQ and IQ scores of Dancer I only differ by country, it can be used as an educational tool for helping people worldwide get on the same page. Dancer I's video could be a good video to show and discuss when judge certifiers trot the globe. Likewise, Dancer D seems to have some aspect of PQ that champion judges see, but others do not, so it can also be used to improve education. Dancers, viewers and prospective judges could all benefit from hearing how those with CJX evaluate Dancer D's PQ. Considering the greater variance of scoring, adjudicator certification can be improved with extra focus on IQ and how it should be interpreted. With IQ seemingly the hardest criteria to score, AQ is seemingly the second hardest, so AQ might require more learning than PQ.

Comparing total score averages of each performance, by method, rankings stayed mostly the same (Table 4). Between Dancers D and E, Dancer E consistently scored higher. If Dancer D was in a battle against Dancer B, Dancer D would have won no matter which judging method was used (personal or Trivium). Dancer G would have lost to Dancer B, and Dancer A would have lost to Dancer G, regardless of the method. Dancers F and G's battle always ends with Dancer G's win. Dancers H and I's battle was very close – personal scoring ends in a tie – but Trivium judging ends with a win for Dancer I. Using Trivium, respondents who scored Dancer A higher than Dancer I suddenly reverse the rankings, as happens also with Dancers C and F.

Table 4: Overall average score comparisons.

Dancer	Personal score	Trivium score
E	4.07	3.96
D	3.74	3.39
B	3.49	3.29
G	3.29	3.13
A	2.68	2.39
I	2.52	2.40
H	2.52	2.11
F	2.34	1.85
C	2.06	1.97

S2 personal judging

S2 did not ask for scores, asking instead to know which side won the battle. The logistic regression model of Battle 1's personal judging results ($N = 109$) found significant correlation between AG, BX, JX, CJX, CT and battle outcome, $X^2(10) = 101.095$, $p < 0.05$. This model explains a large proportion of variance in judging (pseudo $R^2 = 0.604$). 77 respondents chose the left side to win (Dancer E) while only 32 respondents chose the right side to win, but all nine respondents with CJX chose the right side (Dancer D).

Battle 2's personal scores ($N = 109$) found significant correlation between all variables, $X^2(10) = 93.636$, $p < 0.05$. This model has strong predicting power (pseudo $R^2 = 0.576$). 86 respondents chose right and 23 respondents chose left. All championship judges chose right (Dancer G beats Dancer F).

Battle 3's personal scores ($N = 109$) found significant correlation between all variables, $X^2(10) = 46.969$, $p < 0.05$. This model has strong predicting power (pseudo $R^2 = 0.350$). 61 respondents chose left while 48 chose right. This battle was split among championship judges (five chose Dancer H and four chose Dancer I).

S2 Trivium judging

S2 does not ask for scores, instead asking which side had better PQ, IQ and AQ. Models of the PQ decisions for all battles were significant ($p < 0.05$) and had strong predicting power (pseudo $R^2 > 0.39$). Battle 1 was split, with only one more championship judge and a dozen more other respondents deciding that left side (Dancer E) had better PQ. Battle 2 was more decisive, with 29 more respondents, including all championship judges, choosing the right side (Dancer G) as having better PQ. Battle 3 was likewise decisive, with nineteen more people, including all championship judges, choosing Dancer I as having better PQ.

For Battle 1, 85 respondents judged left (Dancer E) as having better IQ, but all nine championship judges chose Dancer D as having better IQ. The other respondents who judged right had average JX = 6, while those who judged left only had average JX = 2 (Figure 6). It was more likely for someone with JX > 5 to interpret this quality of this battle like a championship judge. AQ was more decisive, as 101 respondents, including the championship judges (and even most of those with BX = 0) all chose Dancer E's AQ as higher, indicating that the AQ of this battle was easy to judge.

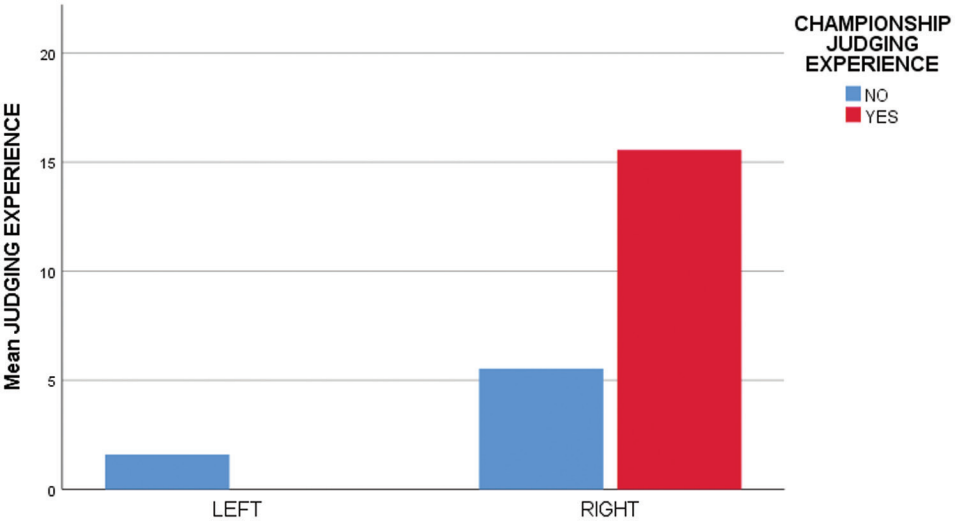

Figure 6: Battle 1's IQ.

Battle 1 is more complicated to judge when choosing sides instead of scoring points. With Trivium, it becomes a question of whether Dancer D's higher IQ rises over Dancer E's higher AQ. The champion judges seem to think so. Such battles are good examples for educational discussion, so that all can better understand how to interpret breaking battles like the experts do.

Battle 1's AQ was split among the championship judges in S1, although Dancer E's S1 total average AQ was higher. Championship judges in S1 scored Dancer E as having higher IQ than Dancer D, but Dancer D's IQ was judged higher in S2. Such discrepancies highlight how useful this video can be as a teaching and learning tool. There might be some bias when thinking about scores that end with different results when judges are thinking about whose IQ is better, and this affects even championship judges. Measuring the variable as scale or as nominal seems able to change battle results.

For Battle 2, 85 respondents judged Dancer G's IQ as higher, including all nine championship judges. Those who judged like the championship judges had an average JX = 19, while the group that chose left had average JX = 8 (Figure 7). Regarding AQ, 94 respondents also judged in favour of Dancer G, including all championship judges. Those with BX = 0 also judged in favour of Dancer G. The few who chose Dancer F's AQ as higher had $3 < BX < 14$, but with random sampling, Dancer G usually gets voted to win against Dancer F, regardless of the method.

For Battle 3, 70 respondents judged Dancer I's IQ as higher. Four championship judges chose Dancer H. Those who chose Dancer I had average JX = 20 and those who chose Dancer H had average JX = 15; however, most respondents with $JX < 7$ chose Dancer I. CT was once again found significant, as most people from Israel, Italy and Holland chose Dancer H while most people from the United States, Germany, South Africa and Canada chose Dancer I (Figure 8). This indicates either a difference in understanding by region or an effect of uneven sample sizes, but the former is probably true due to how not one of the 24 South African respondents chose left.

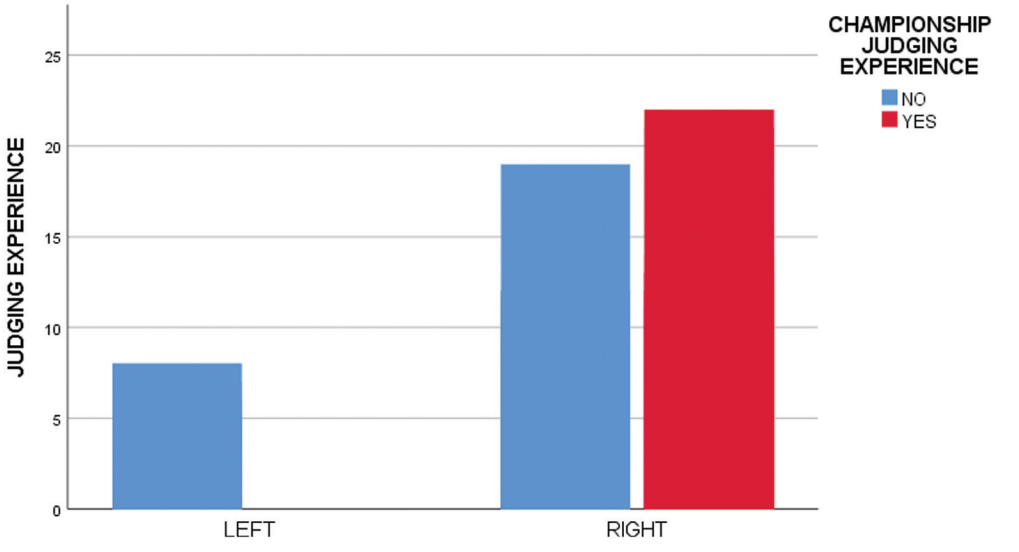

Figure 7: Battle 2's IQ.

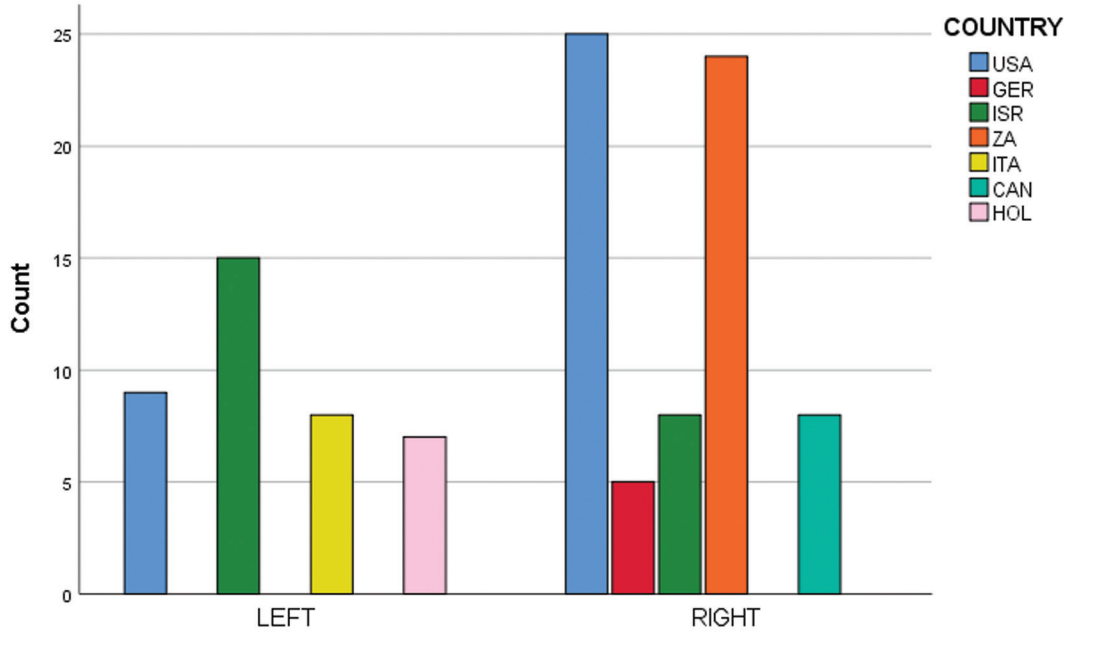

Figure 8: Battle 3's IQ.

AQ was likewise split. 72 respondents judged for Dancer H. Five championship judges voted left while four voted right. Most respondents with little or zero BX also voted left, along with all South African respondents. Battle 3 seems like it can go either way depending on who is judging and how.

Comparing all outcomes and considering variance, the three b-girls (C, H and I) are most anomalous. Effects of sample size are unlikely because no such influence was found for the b-boy scores. The b-girl performances probably included elements that are difficult to uniformly interpret. While CJX or evaluation method can explain variance in b-boy results, the same do not explain variance in the b-girl results. Sometimes, it was the effect of AG, of CT, of BX, of JX or of CJX, and sometimes it was a mix of many or all effects that linked to respondents' differing views as to how well the b-girls did, who should win the b-girls' battle, and who scores higher for which qualities. Even among the most veteran judges and the championship judges, the b-girls' battle was split by IQ and AQ. Disconnected from veterancy, Battle 3 shows judgement decisions split only by CT, with most and sometimes all people from certain countries making the same choice regarding which b-girl had the better IQ and AQ. The femininity of Dancers C, H and I possibly prompted different and disorganized assessment, but it could also be coincidence – that these three dancers just happen to practise unique approaches to breaking that have nothing to do with gender.

DISCUSSION

This study explored the relationships between country, veterancy and the evaluation of dancer performances through different lenses, such as through the Trivium lens. It was found that age, dance veterancy, judge veterancy and country can all influence how one judges dancers' solo and battle performances. Analysis results imply that IQ and AQ are more difficult to judge than PQ. It also seems that the older and more experienced a judge is, the more that judge is likely to score performances similarly to those with CJX. On the other hand, some battles and some dancer qualities are easily understood even by dancers with fewer years of experience, while even the championship judges give split decisions on some battles and dancer qualities. The b-girls' battle, for example, resulted in such a split in which CT, curiously, was highlighted as a potentially very important shaper of a viewer's interpretation of what the b-girls were dancing.

Gender is often studied in the field of breaking, usually revealing instances of bias and inequality (Gunn 2016, 2017, 2021). Breaking is perceived by many as a male culture and, in their eyes, a female exists as an 'other' within it. Breaking can be an androgynous and non-heteronormative space, but distinctions are often raised by both sexes, as female participants might also call attention to their distinguished identities. For example, a video of female dancers at a breaking competition produced by b-girls chose to highlight their femininity, as a way of balancing between masculine and feminine energies at the event (Gunn 2021). Women have been breaking since the beginning of the dance's history (Aprahamian 2020) and breaking can benefit women in multiple developmental ways (Gunn 2016). The culture has potential to be an inclusive space for women to thrive in (Gunn 2017), but regarding battle adjudication, the backgrounds of people hailing from different countries can influence their judgement of female breaking performances.

Dance and judging veterancy were also found in both S1 and S2 to be frequently influential factors, but this issue might be harder to tackle than gender. Attitudes concerning non-male breakers, seemingly a matter of nationality, can be attended to by providing customized judge education per country. Regarding veterancy, even people with none at all can sometimes score like champion judges. At other times, even championship judges can be divided in their evaluations. With so many factors involved, a close battle with little skill imbalance can be a challenge for even the most experienced veterans to judge. Likewise, a battle between dancers who both excel at a specific yet different aspect of breaking can be just as challenging. For these reasons, judge certification vetting need not discriminate.

If more applicants were invited to try learning the system and passing the test, the World DanceSport Federation could earn more judges, and also earn more money from the influx of application fees. In this case, the exam would be the threshold for veterancy and it would become the most important filter of judge quality. As such, more study should be done on the exam itself, and the expertise of judges who pass it. The certification course could also be improved by attending to the variance in quality scores.

Teachers of the certification course can prepare materials and conversate more on IQ and AQ, which seem to be less widely understood than PQ. Issues of judging close battles can be explained by veteran judges who provide examples for their decisions and the reasons behind them. Videos, such as the video of Dancer D, can be used as examples for the champions to explain what it is that they see that most others seem to not yet notice. As a curriculum preparation tool, this survey and response analysis method can be useful for choosing which videos to bring and what aspects of them to focus on.

This study has shown how some videos can be useful for judge education due to their various merits. Videos can be analysed using surveys, as done here, to understand the details of the variance of their scoring and use them as examples to dissect within judge training workshops. This tutelage extends out to the general public, as promoters of Olympic breaking can come to recognize what the event audience do not understand about competitions and their adjudication, then provide relevant education to the battle audience.

With a contestants' gender possibly being a factor in battle outcomes, a judge's gender could also perhaps be a factor. Potential other variables that influence battle judging can include dance style, judge alertness and even cognitive load (thinking about numbers, such as in S1, can get different results than when thinking about sides, as in S2). Judging style can also make a difference, if some tend to prioritize their overall impression while others tend to think about the smaller details of each quality or perhaps only pay attention to a few highlights of each round. Despite all judges supposedly using the same system and guidelines to judge Olympic breaking events, the realities of judging in practise seems to not be so uniform.

Research on battle judging has only begun to explore its depth. Future research can break through the limitations of this current work by considering the respondent's gender, defining some battles as more difficult to evaluate and running analyses with such difficulty variable also in the model, and also trying before–after surveys of those who passed the Trivium certification exam, to account for more variables and to measure the effect of the education provided. Studies should strive to continue using larger sample sizes, but quality is as important as quantity. With a largely dispersed survey, researchers

cannot know if the anonymous respondent really has ten years of CJX. Surveys can be sent directly to certain members of the intended sample populations to generate more accurate insights. It should be important to also control for differences in knowledge about the Trivium system and its use, such as by asking people if they read the manuals, tried using the system or watched videos explaining its criteria. Instead of using Likert scales or binary items, future surveys could become more valid if they used the actual Trivium system itself. Such studies should be done on all WDSF and the Olympic breaking judging systems (such as the newer Level A and Level B systems). The teaching practice methods used during judge certification can also be researched to find ways of improvement, and qualitative methods can delve deeper to discover how people of varying veterancy, and from different regions, understand the breaking that they see.

CONCLUSION

It was expected that people from different countries, as they become more experienced dancers and judges, eventually reach similar understandings of how to interpret the battles they see. Evidence was found to support the claim that older and more experienced dancers and judges are often more likely to decide as the champion judges do (research question 1), but it might be more a matter of the dance/battle itself. For some performances, even the champion judges are split, and for some very close competitions or difficult to interpret qualities, veterans can lack the championship judge eye no matter how much experience they have. When evaluating b-girls, judges' nationality can significantly vary their scores (research question 2). In all, it seems as though it is more difficult to judge IQ and AQ than PQ (research question 3). While preparing for Olympic competition judging, those appointed should practise judging until their b-boy and b-girl evaluations become more consistent and impartial, accompanied by teachers who can help them better understand what the founders of the Trivium system meant when they spoke of AQ and IQ.

Although the models tested here were found to have strong predicting power, the tested predictors could be overshadowed by the nature of each battle and the education of each viewer. Those with few years of dance experience can learn to get on the same page, but some battles will always be difficult to evaluate and return split decisions. Considering this, the certification courses and their final exams become the main definers of who is a good judge, so future studies should be aimed at improving these courses and exams. The findings we presented can help advance breaking education for judges, contestants and Olympic viewers as this dance prepares to step into its new spotlight.

REFERENCES

Anderlucci, L., Lubisco, A. and Mignani, S. (2021), 'Investigating the judges performance in a national competition of sport dance', *Social Indicators Research*, 156:2, pp. 783–99.

Aprahamian, S. (2020), '"There were females that danced too": Uncovering the role of women in breaking history', *Dance Research Journal*, 52:2, pp. 41–58.

Baker, G. (2012), 'Mala Bizta Sochal Klu: Underground, alternative and commercial in Havana hip hop', *Popular Music*, 31:1, pp. 1–24.

Bryman, A. (2012), *Social Research Methods*, 4th ed., Oxford: Oxford University Press.

Dalecki, L. (2011), 'Mining the cypher: Branding and breakdance', *International Journal of Culture, Tourism and Hospitality Research*, 5:4, pp. 364–75.

Edmonds, W. A. and Kennedy, T. D. (2016), *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*, Los Angeles, CA: Sage Publications.

Fahr, D., Gopie, K., Graham, S., Grigo, P., Ma, G., Robitzky, N., Roth, M. and Shaughnessy, T. (2018), *Buenos Aires 2018 Youth Olympic Games WDSF Rules and Regulations Manual: Breaking*, Lausanne, CH: World DanceSport Federation.

Fogarty, M. (2019), 'Why are breaking battles judged?', in S. Dodd (ed.), *The Oxford Handbook of Dance and Competition*, Oxford: Oxford University Press, pp. 409–29.

Fowler Jr, F. J. (2013), *Survey Research Methods*, Thousand Oaks, CA: Sage Publications.

Gliem, J. A. and Gliem, R. R. (2003), 'Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales', *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*, 8–10 October, Columbus, OH: Ohio State University, https://scholarworks.indianapolis.iu.edu/server/api/core/bitstreams/976cec6a-914f-4e49-84b2-f658d5b26ff9/content. Accessed 12 June 2024.

Gunn, R. (2016), 'Don't worry, it's just a girl! Negotiating and challenging gendered assumptions in Sydney's breakdancing scene', *Journal of World Popular Music*, 3:1, pp. 54–74.

Gunn, R. (2017), 'Deterritorializing gender in Sydney's breakdancing scene: A b-girl's experience of b-boying', Ph.D. dissertation, Sydney: Macquarie University.

Gunn, R. (2021), 'Where the #bgirls at? Politics of (in) visibility in breaking culture', *Feminist Media Studies*, 22:6, pp. 1447–62.

Haase, R. F. (2011), *Multivariate General Linear Models*, Los Angeles, CA: Sage Publications.

Harrell, F. E. (2015), 'Binary logistic regression', in *Regression Modeling Strategies*, New York, NY: Springer, pp. 219–74.

Li, R. Z. and Vexler, Y. A. (2019), 'Breaking for gold: Another crossroads in the divergent history of this dance', *The International Journal of the History of Sport*, 36:4&5, pp. 430–48.

Sato, N. and Hopper, L. S. (2021), 'Judges' evaluation reliability changes between identifiable and anonymous performance of hip-hop dance movements', *Plos One*, 16:1, https://doi.org/10.1371/journal.pone.0245861.

Shaughnessy, J. J., Zechmeister, E. B. and Zechmeister, J. S. (2000), *Research Methods in Psychology*, New York: McGraw-Hill.

Shimizu, D. and Okada, T. (2018), 'How do creative experts practice new skills? Exploratory practice in breakdancers', *Cognitive Science*, 42:7, pp. 2364–96.

VanderWeele, T. J. and Mathur, M. B. (2019), 'Some desirable properties of the Bonferroni correction: Is the Bonferroni correction really so bad?', *American Journal of Epidemiology*, 188:3, pp. 617–18.

Vexler, Y. A. (2021), 'Breaking for gold: Dance education & knowledge integration', MA thesis, Jerusalem: Hebrew University of Jerusalem, https://huji.userservices.exlibrisgroup.com/discovery/delivery/972HUJI_INST:972HUJI_V1/12319580470003701?lang=en&viewerServiceCode=AlmaViewer. Accessed 2 January 2022.

Yang, Z., Bai, Y. and Wei, M. (2022), 'The importance of creativity in the sportification of breakdance', *Frontiers in Education*, 7, pp. 1–5, https://doi.org/10.3389/feduc.2022.855724.

SUGGESTED CITATION

Vexler, Yonatan Asher, Foelix, Candy and Foley, Cindy (2023), 'Breaking for gold: Who can judge me?', *Global Hip Hop Studies*, Special Issue: 'Breaking and the Olympics 2', 4:2, pp. 127–48, https://doi.org/10.1386/ghhs_00095_1

CONTRIBUTOR DETAILS

Yonatan Asher Vexler is a postgraduate researcher with over twenty-five years of experience in breaking, also as a teacher and professional translator. His research focuses on education, training, culture, curricula and sociology, along with other issues concerning sport, dance, martial arts and even language learning. He holds all five World DanceSport Federation (WDSF) breaking adjudication certificates (including chair/head judge), multiple coaching and teaching certificates, and he is also a thirteenth-generation disciple of Chen-shi taijiquan and a 34th-generation disciple of Shaolin. Always open to scientific collaboration, feel free to e-mail him at any time. He is currently involved with a new project investigating youth's motivations to learn in general and practice breaking in particular, but would gladly help with other projects, too. His worldwide b-boy journey includes battling in America, cyphering in Europe, judging in Asia, teaching in Africa, performing on stages, organizing events, interviewing veteran breakers, and even trying to promote peace and love by breaking in the Middle East.

Contact: Child of this Culture Foundation, 3121 Pell Mell Drive, Orlando, FL 32818, USA.
E-mail: yoni.vexler@mail.huji.ac.il

https://orcid.org/0000-0002-1199-6259

A simple T-shirt idea turned twin sisters, Candy Foelix and Cindy Foley's passion for hip hop into a leadership opportunity. Over a decade ago, they founded the Child of this Culture Foundation (COTC) based in Orlando, Florida, United States. COTC is a community-based organization that supports the development of breaking into the United States, protects the image of hip hop in educational systems, and empowers the next generation to find a voice through dance. In 2018, because of breaking's debut at the Youth Olympic Games in Buenos Aires, the expansion of the Hip Hop Arts, Culture, and Sport Placemaking Strategic Plan formed the first idea to develop a strategic plan to assess the agency's creative placemaking efforts to date and support research and development. Candy Foelix, president of COTC, works as an intuitive trauma coach, bodyworker, wife and mother of two with a hip hop cultural background. Her specialty is aiding people in healing through neuromuscular therapy, somatic therapy and myofascial release. She also contributes to BMJ Open Sport & Exercise Medicine and is a freelance commentator for the Olympic Channel.

Contact: Child of this Culture Foundation, 3121 Pell Mell Drive, Orlando, FL 32818, USA.
E-mail: childofthisculture@gmail.com

https://orcid.org/0009-0000-7590-9276

Twin sisters Candy Foelix and Cindy Foley founded the Child of this Culture Foundation (COTC) over a decade ago. COTC is a community-based organization that supports the development of breaking into the United States, protects the image of hip hop in educational systems, and empowers the next generation to find a voice through dance. In 2018, because of breaking's debut at the Youth Olympic Games in Buenos Aires, the expansion of the Hip Hop Arts, Culture, and Sport Placemaking Strategic Plan formed the first idea to develop a strategic plan to assess the agency's creative placemaking efforts to date and support research and development. Cindy Foley is a veteran and military spouse pursuing a doctorate at the Rossier School of Education at the University of Southern California. She has published an article called 'From bgirl to bbaby: Constructing a breaking pedagogy' in the *Journal of Dance Education* (2016). She serves COTC as executive director.

Contact: Child of this Culture Foundation, 3121 Pell Mell Drive, Orlando, FL 32818, USA.
E-mail: childofthisculture@gmail.com