



“It’s Always the Judge’s Fault”: Attention, Emotion Recognition, and Expertise in Rhythmic Gymnastics Assessment

Lindsey G. van Bokhorst¹, Lenka Knapová^{2*}, Kim Majoranc³, Zea K. Szebeni⁴,
Adam Táborský², Dragana Tomić⁵ and Elena Cañadas⁶

¹ Maastricht University, Maastricht, Netherlands, ² Masaryk University, Brno, Czech Republic, ³ University of Ljubljana, Ljubljana, Slovenia, ⁴ Eötvös Loránd University, Budapest, Hungary, ⁵ University of Banja Luka, Banja Luka, Bosnia and Herzegovina, ⁶ University of Lausanne, Lausanne, Switzerland

OPEN ACCESS

Edited by:

Agnieszka Walczak,
Cambridge English Language
Assessment, UK

Reviewed by:

Antonio Calcagni,
Hungarian Academy of Science,
Hungary

Chris J. Gibbons,
University of Cambridge, UK

*Correspondence:

Lenka Knapová
knapova.l@gmail.com

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 08 March 2016

Accepted: 20 June 2016

Published: 05 July 2016

Citation:

van Bokhorst LG, Knapová L,
Majoranc K, Szebeni ZK,
Táborský A, Tomić D and Cañadas E
(2016) “It’s Always the Judge’s Fault”:
Attention, Emotion Recognition,
and Expertise in Rhythmic
Gymnastics Assessment.
Front. Psychol. 7:1008.
doi: 10.3389/fpsyg.2016.01008

In many sports, such as figure skating or gymnastics, the outcome of a performance does not rely exclusively on objective measurements, but on more subjective cues. Judges need high attentional capacities to process visual information and overcome fatigue. Also their emotion recognition abilities might have an effect in detecting errors and making a more accurate assessment. Moreover, the scoring given by judges could be also influenced by their level of expertise. This study aims to assess how rhythmic gymnastics judges’ emotion recognition and attentional abilities influence accuracy of performance assessment. Data will be collected from rhythmic gymnastics judges and coaches at different international levels. This study will employ an online questionnaire consisting on an emotion recognition test and attentional test. Participants’ task is to watch a set of videotaped rhythmic gymnastics performances and evaluate them on the artistic and execution components of performance. Their scoring will be compared with the official scores given at the competition the video was taken from to measure the accuracy of the participants’ evaluations. The proposed research represents an interdisciplinary approach that integrates cognitive and sport psychology within experimental and applied contexts. The current study advances the theoretical understanding of how emotional and attentional aspects affect the evaluation of sport performance. The results will provide valuable evidence on the direction and strength of the relationship between the above-mentioned factors and the accuracy of sport performance evaluation. Importantly, practical implications might be drawn from this study. Intervention programs directed at improving the accuracy of judges could be created based on the understanding of how emotion recognition and attentional abilities are related to the accuracy of performance assessment.

Keywords: rhythmic gymnastics, attention, emotion recognition, expertise, accuracy, judges

INTRODUCTION

Judges are often in the center of media coverage, repeatedly criticized, and sometimes undervalued part of the sport (Dosseville et al., 2014). Research concerning judges mostly focuses on the differences between novice and expert judges (Ste-Marie, 1999, 2000; Flessas et al., 2015) and the biases that occur during competitions (Ansorge and Scheer, 1988; Plessner, 1999;

Plessner and Schallies, 2005). However, no attention has been devoted to psychological factors such as attention, emotion recognition of judges and possible interventions which could reduce biases or stress. Mentioned factors are relevant and may influence the outcome in sports with subjective scoring, such as rhythmic gymnastics, where the results – scoring and ranking of performances – depend heavily on evaluations made by judges.

Typically the scoring, as a main tool of evaluation, used in rhythmic gymnastics is based on subjective decisions of the judges. These decisions might be affected by cognitive abilities of judges such as selective attention and vigilance. Lower cognitive abilities affect the processing of visual information (Glisky, 2007). During a competition, judges have to track multiple objects at the same time which requires high demands on their resources. Research examined that individuals can identify multiple objects simultaneously, and in case of professional judges, the number is significantly higher (Cavanagh and Alvarez, 2005). It seems training and expertise increase cognitive resources available to process visual information. Oksama and Hyönä (2004) showed that the capability for tracking multiple objects can be affected by factors such as increased speed of the objects (which decreases capabilities) and the novelty of the objects (familiar objects increase capabilities).

Throughout the evaluation of rhythmic gymnastics performances, the judges have to follow the performance of five or six gymnasts at once, while each of the gymnasts also holds an apparatus (hoops, balls, or ribbons) which becomes separated from them during throws and other elements. While paying attention to bodily movements, the skilful movement of the apparatus, coordination, etc., judges face several distracting factors such as the color of the gymnasts' dresses (Hagemann et al., 2008), the size of the crowd around them (Bai et al., 2007), the noise (Nevill et al., 2002), the disadvantageous angle from which they sometimes have to observe the performance (Oudejans et al., 2000), and even some psychological factors such as confirmation bias (Rodenberg, 2011).

Even though, rhythmic gymnastics settles rare conditions, it includes some human aspects which might be part of an evaluation and which cannot be avoided as emotions are present in every context of social interaction, including sport performance. They play an important part in the artistic evaluation of the performance, since the body movements and facial expressions have to be harmonized with the music. Emotion recognition is the ability to identify emotions of others through their nonverbal expression including face and body (Gabrielsson and Juslin, 1996). For gymnasts, emotions can influence sporting behaviors and actions through their effect on attention, and therefore concentration (Lazarus, 2000). Vast et al. (2010) studied the effect of perceived emotions on attention, especially concentration, and performance. They found that anger negatively affects concentration but not performance, whereas anxiety negatively affects performance but not concentration. Emotion recognition might play an important role during the evaluation of the performance in two ways. First, being able to accurately discriminate emotional intentions and expression of the gymnasts can improve assessment of the performance due to the amount of information that emotions

provide about the situation and the mental state of the gymnasts. Second, emotions play an important role in error detection. For instance, a brief expression of anger or disappointment by the gymnasts can be an alerting signal to judges that a mistake was committed. Without emotion expression, the mistake could have passed unnoticed.

Not only emotions play an important role in evaluation. The expertise of judges has been found to influence the evaluation of performance of gymnasts. Ste-Marie (2000) found that novice judges spend less time looking at the gymnasts and more time looking at the scoring paper than the judges with more expertise. Flessas et al. (2015) confirmed that expert judges have better error detection than novices. However, even the highest ranked international judges only reported 40% of true errors. Other research suggests that expert judges are better at perceptually anticipating upcoming gymnastic elements, which has a positive effect on their judging performance as compared to novice judges (Ste-Marie, 1999). A study, done with football referees, obtained positive correlation indices among hours practiced per week, the number of competitions judged, and the ability to evaluate a performance (Catteuw et al., 2009). Studies have also shown that both visual and motor experience with a specific sport account for correctly estimating the movement quality of other people (Loula et al., 2005; Blake and Shiffrar, 2007). However, according to a study concerning practical skills rhythmic gymnastics judges and their relation to the judging abilities was founded that judges mostly valued following skills: the knowledge of technical parameters of the sport and the capacity to adjust to any level of competition under self-assuredness and self-confidence circumstances, but not experience *per se* (Fernandez-Villarino et al., 2013).

Another factor that can greatly affect rhythmic gymnastics evaluation is chronotype of the judges. Chronotype is known as a trait that reflects individually preferred times for activity and sleep (Smith et al., 1989; Greenwood, 1994; Caci et al., 2000). According to the preferred times for activity, chronotypes are classified into “morningness”, “intermediate”, and “eveningness” types. Individuals of the morningness type prefer to be active in the early morning, whereas those of the eveningness type prefer to be active late in the evening. Chronotype is related to attention peaks at different times of day. Previous studies have shown that morning-types are more alert in the morning (Clarisse et al., 2010), while evening-types' alertness did not differ between morning and afternoon (Adan and Guàrdia, 1993) or they performed better in the evening (Schmidt et al., 2012). However, some studies found no association between chronotype and performance inattention tests (Adan, 1991; Gomes et al., 2011; Adan et al., 2012).

Due to mentioned factors it might be beneficial to use the Rasch model as it has been previously reported to be beneficial in assessing sport performances, and more concretely to evaluate consistency among the different evaluations of a judge and the different aspects of performance been evaluated (Looney, 2003).

Previous research on judged scoring in sports has mostly focused on the differences between novice and expert judges, but research on psychological factors (i.e., attention and emotion recognition) and their influence on performance assessment

have yet to be investigated. The present study investigates how attentional and emotion recognition abilities can affect accuracy of sport assessment in rhythmic gymnastics. The primary goal of this study is to evaluate attention and emotion recognition in rhythmic gymnastics judges and their relationship with accuracy in performance evaluation. Furthermore, the aim is to investigate possible differences between novice and expert judges. Finally, the results could help in improving the accuracy and future training of judges. We hypothesize that (i) attentional and emotion recognition abilities of judges are positively correlated with scoring accuracy, (ii) higher level of expertise is positively correlated with scoring accuracy, and (iii) expertise has a mediating role between attentional and emotion recognition abilities of the judges and scoring accuracy.

MATERIALS AND METHODS

Materials and Measures/Equipment Video

Participants will evaluate a set of videotaped group rhythmic gymnastics performances (2–3 min each) on the artistic and execution components of the performance score. Using recordings of past performances will ensure the same conditions for all participants and that will allow us to compare the differences in scoring between participants. Detailed information about the selected videos can be found in **Table 1**. These videos were selected in a two-step process. Firstly, all the videos from past Olympic Games freely available online were gathered. The Olympic Games were chosen for its best video coverage as well as broadest competition. It was decided that videos from the latest 2012 Olympic Games should be used with respect to the recording quality that ensures better visibility of performance details that is crucial for accurate performance evaluation. In the second step, the videos were studied regarding the obtained score, noticeability of errors, and dress conspicuousness. In the end, four videos including a range of performances from almost perfect performances to performances with more noticeable errors were selected. The resolution of the videos is 720p and the sound is muted to prevent the participants' evaluations from being influenced by the commentary.

A certain risk to the study is the fact that judges might remember the scores of the performances that were given at the Olympic Games. To check for this, the question: "Have you seen this performance before?" will be asked after each video. If their answer is yes, then they will be asked to write down the remembered scores. This will allow us to see whether they actually

remembered the score or whether the performance just looked familiar to them and control for this in the analysis.

Attentional Abilities

The attentional abilities will be measured by the Attention Network Test for Interactions and Vigilance (ANTI-V; Roca et al., 2011; including new measures of vigilance executive and vigilance of activation – ANTI-VEA) that takes about 15 minutes to complete. The test (**Figure 1**) is computer-based and measures participants' performance in five components of attention: phasic alertness, vigilance executive, vigilance of activation, orienting network, and executive control. ANTI-VEA measures these components through reaction time and accuracy (or percentage of errors). Participants have to indicate the direction of the central stimulus, by pressing C or M on the keyboard (typical flanker task). Direction of the flanker stimuli (either congruent or incongruent) tests the executive control functioning. When participants detect that the central stimulus is displaced, they have to press spacebar. This type of stimuli is unpredictable, infrequent, and unexpected and detection of such a stimulus measures vigilance executive, as Interactions and Vigilance (ANTI-V; Roca et al., 2011). Vigilance of activation is measured by the Psychomotor Vigilance Task (PVT; Correa et al., 2014), a numerical count down starting at 999 ms that participants have to stop as fast as possible by pressing any key. Visual cue assesses the functioning of the orienting network (valid, invalid, or no cue condition). Auditory warning signal (50 ms) that announces the appearance of the stimulus measures phasic alertness.

The advantage of ANTI-VEA is that it tests the five attentional components in an independent way and allows measuring of the interactions between them. Moreover, it measures the reaction time for every answer. Reliability of ANTI-V was 0.99 (Roca, 2012) and we expect the ANTI-VEA to be highly reliable as well as it is a new, slightly adjusted, version of ANTI-V.

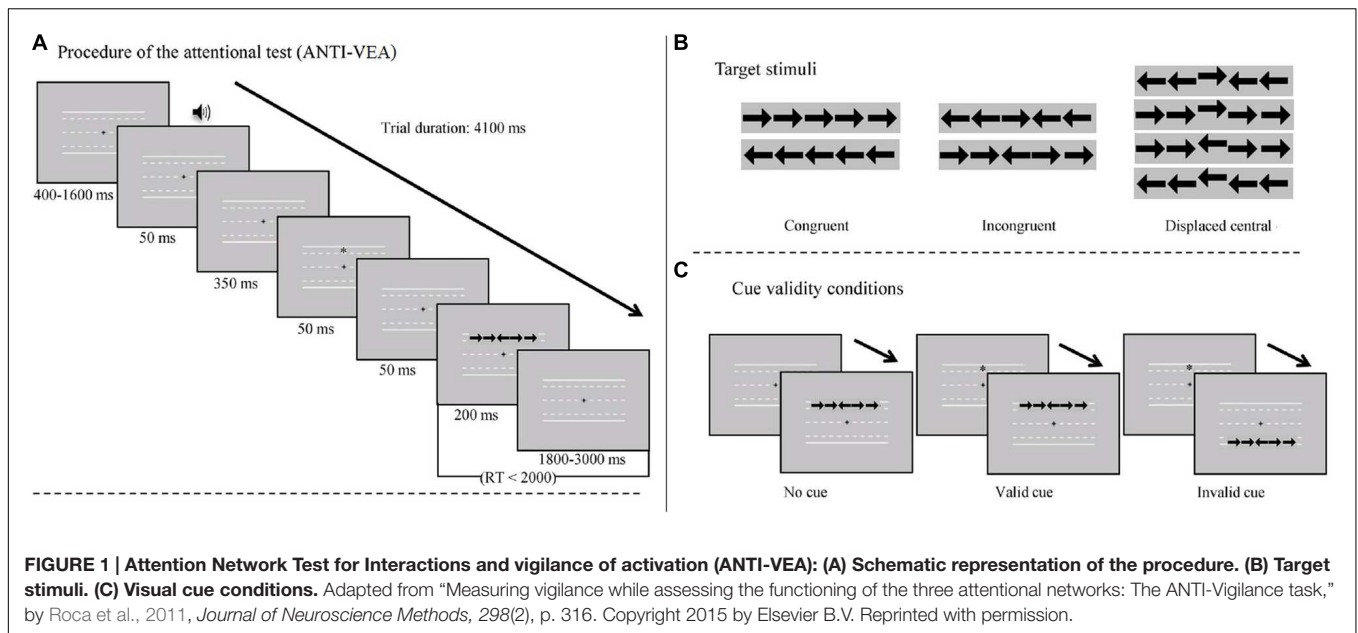
Emotion Recognition Ability

The Geneva Emotion Recognition Test-Short (GERT-S; Schlegel and Scherer, 2015) measures individual differences in the ability to recognize emotions of others. It is a shortened version of the Geneva Emotion Recognition Test (GERT; Schlegel et al., 2014). GERT-S consists of 42 items short audio-visual clips in which ten actors express 14 different emotions (e.g., anger, relief, joy, pride, and anxiety) through their face, voice, and body. After each clip, participants have to choose one of the 14 emotions which they believe was expressed by the actor. This test is dynamic and multimodal (short video

TABLE 1 | Chosen group rhythmic gymnastics performances from 2012 Olympic Games.

Competition	Discipline	Nation	Score artistic	Score execution	Place
Qualification	3 hoops, 2 ribbons	Russian Federation	9.550	9.100	1st
Qualification	3 hoops, 2 ribbons	Greece	8.900	8.450	9th
Final	3 hoops, 2 ribbons	Spain	9.300	8.800	4th
Final ^a	3 hoops, 2 ribbons	Bulgaria	9.450	8.850	6th

^aThis video will be used as a practice video to get familiar with the judging process.



clips with sound) and measures ERA in a more ecologically valid fashion. In comparison with other ERA tests which usually offer basic emotions, GERT-S features 14 different emotions, both positive and negative. Moreover, GERT-S internal consistency is between 0.80 and 0.83 (Schlegel and Scherer, 2015).

Expertise

Participants will be asked about their judging experience (judging category, years of judging at the international level, and number of judged competitions) to determine their expertise. There are four categories of FIG judges for all disciplines, Category IV being the lowest and Category I the highest. Category IV Judges are new international judges with no or little international experience. They are not assigned as judges in major competitions. Category III Judges are defined as experienced judges with good results in execution and artistic. They are designated to judge execution and artistic in major competitions. Category II Judges are experienced judges with very good results in difficulty. They are designated to judge difficulty as well as execution and artistic. Category I Judges are very experienced judges with excellency in difficulty and operate as members of the Superior Jury, Chair of Judges' Panel, and Difficulty Judges and are also designated to judge execution and artistry (Fédération Internationale de Gymnastique [FIG], 2013).

Additional Questions

Participants will further answer basic demographic questions (gender, age, nationality, etc.), questions about their chronotype (rMEQ; Adan and Almirall, 1991; example question: *At what time in the evening do you feel tired and as a result in need of sleep?*), and manipulation check questions (whether they have seen the presented performances before and whether they remember the scores).

STEPWISE PROCEDURE

Power Analysis

The software G*power (Faul et al., 2007) was used to calculate the required sample size. Values for α were set on 0.05 and power on 0.80. Based on previous literature and discussions between the authors, effect sizes were estimated on 0.15 (medium effect). In total, to reach the desired power, data from 55 participants has to be collected.

Qualtrics Survey

The study uses an online survey designed on Qualtrics which enables judges from different countries to participate. There is a great benefit of using videotaped performances that can be assessed on any computer over real-time evaluations at competitions. With regards to the different nationalities of the participants, the instructions will be translated into several languages to optimize the comprehensibility of the survey for judges from different cultural backgrounds and native languages. Although translating instructions into different languages might pose limitations to the internal validity of the study, the current study will use native speakers with a fluency in English to translate instructions into their native languages. Participants recruitment and the overall small research population can also cause a limitation in regards to the sample size of the research, but it is important to notice that the actual population of judges in rhythmic gymnastics at international level is also limited (376 judges approved by the FIG). Therefore, it would be interesting in future experiments to include judges from different disciplines, such as dance sport, to extent the research population. Finally, the current study applied a new approach to measure accuracy of performance scoring, as to the authors' knowledge such measuring tools have not been developed nor validated. However, the official scores given at international competitions

are considered as the most accurate attainable scores, and it can therefore be concluded that participants that gave similar scores have high scoring accuracy.

Participant Recruitment

This study aims to recruit 50 rhythmic gymnastics judges approved by the International Gymnastics Federation (13% of the total number of judges at this level) and a control group of 20 rhythmic gymnastics coaches unskilled in performance judging. Participants will be recruited via email, social media, personal contact, and with the help of national gymnastics federations. After confirming participation via a consent form following the ethical committee approval from the University of Lausanne, they will complete an online testing session. Participation will be anonymous and voluntary.

Procedure

Data will be collected using an online survey design based on Qualtrics including measures of personal characteristics and the assessment of videotaped performances of rhythmic gymnasts. The survey will have multiple language versions. The first part of the survey will focus on the evaluation of group rhythmic gymnastics performances. Initially, participants will be asked to read the International Gymnastics Federation's Code of Points on the artistic and execution components. Participants will then be given one practice video to get familiar with the judging process. After the practice video, participants will be shown three video clips of performances from the 2012 Olympic Games. Participants will watch the videos and score them according to the artistic and execution components. Their scores will be compared to the official scores given at the competition the videos were taken from to determine the scoring accuracy of each participant. The official scores given at the Olympic Games are considered to be the most accurate scores that are attainable for this purpose. The second part of the online survey will consist of two tests measuring attentional (ANTI-VEA) and emotion recognition abilities (GERT-S). Finally, participants will answer additional questions about their judging experience (expertise), chronotype (rMEQ), demographic data, and manipulation check.

STATISTICAL ANALYSIS

To test our hypothesis we will use statistical software including SPSS, R and Stata. We will follow the four-faceted Rasch model (gymnastic group ability, aspect difficulty, program difficulty, and judge severity) to assure that valid evaluation occurred between videos and within judges. To do that, we will follow the protocol recommended by Linacre (1999) and Wright and Mok (2000) by mean of the Mififac Rasch Software. Once confirmed that the data fit the model well enough so judges' performance evaluations are valid, we then will create a variable that indicates how much actual judges' evaluations deviate from Olympic judges' evaluations (accuracy).

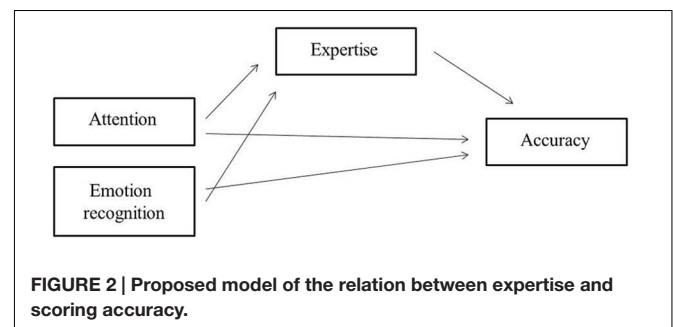
Next, we will perform correlational analysis to evaluate the hypothesized positive correlations between attentional abilities

and accuracy, ERA and accuracy as well as expertise and accuracy. When testing a more complex model that includes causality and interrelations between the variables, we encounter the problem of working with endogenous variables, in our case latent variables that are being estimated through their manifestations, i.e., attention and emotion recognition (Antonakis et al., 2010). Moreover, the relationship between these variables and the dependent variable (accuracy) is expected to partly function through other variables in the model (i.e., expertise). Therefore the data will be further analyzed using regression analysis and Structural Equation Modelling (SEM) to test a mediational role of the relationship between expertise and accuracy (**Figure 2**).

The evaluation of the fit of the model to the obtained data we will rely on the Root Mean Square Error of Approximation index (RMSEA; Steiger and Lind, 1980) that is based on the χ^2 value of the model as well as the number of observed cases and degrees of freedom, as depicted in Equation 1. RMSEA's values of 0.05 and lower indicate a good fit, while values up to 0.08 indicate an acceptable fit (Browne and Cudeck, 1993). We will use also *SWAIN* correction, as a measure to address sensitivities of χ^2 statistic to sample size and model complexity (Antonakis and Bastardo, 2013). RMSEA formula retrieved from <http://davidakenny.net/cm/fit.htm>,

$$\frac{\sqrt{(\chi^2 - df)}}{\sqrt{[df(N - 1)]}} \quad (1)$$

To address the above-mentioned issue of endogeneity and to deliver consistent parameter estimates, the analysis will follow Antonakis et al. (2010) recommendations: (a) include instrumental variables (variables uncorrelated with the error term of the dependent variable and correlated with the endogenous variables) that are necessary to provide a sufficient identification of the model, for instance, age, gender, etc., (b) be performed through the Two-Stage Least-Squares procedure (2SLS; James and Singh, 1978; Baltagi, 2011), a tool used in SEM to estimate path coefficients. In a first-stage, attention and ERA will be regressed on two instrumental variables and covariates, including fixed effects. In the second-stage, accuracy will be regressed on the predicted value of attention and ERA from the first stage and covariates, including fixed effects. We will verify that the instrumental variables used are supported by theoretical and empirical considerations.



ANTICIPATED RESULTS

We expect to support our hypotheses by testing the model presented in the methods section. The first hypothesis stated that attention and emotion recognition are positively related to scoring accuracy. It is assumed that higher levels of attention improve error detection of judges and therefore enable them to rate the performances more accurately, and higher levels of emotion recognition provide judges with the skill to detect changes in facial or bodily expressions that accompany mistakes in the performance. The second and third hypotheses propose a relationship between expertise and attention and emotion recognition and a relationship between expertise and scoring accuracy. By testing our anticipated model we expect to find that expertise is also associated with attention and emotion recognition, and expertise itself is associated with scoring accuracy. Better expertise, defined both as level and years of judging, provides judges with more experience and training to judge more accurately, and it may therefore be assumed that judges of higher levels would be significantly more accurate in their assessments.

The study is facing some potential pitfalls when collecting data. For example, as participants include judges at international levels, some may have already seen the performances to be

evaluated in the survey and therefore remember the scores that were given at the competition. We will control for positive answer to those questions. Another important factor that is included in the survey is whether participants are morning or evening types (i.e., chronotype), as different types reach their peak level of attention and concentration at different moments during the day. Not completing the survey within one's peak hours may possibly influence the person's accuracy and scores of the ANTI-VEA negatively. To take this into account, we control for both self-reported chronotype as well as the moment of the day the tasks were performed, to make sure everyone was at their full activity range.

AUTHOR CONTRIBUTIONS

EC is the author of the concept of the study, the supervisor of the project, and provided feedback when writing the manuscript. KM, ZS, and AT were in charge of writing the introductory part of the manuscript (including literature review). LvB, LK, and DT were in charge of writing the Materials and Methods, Stepwise Procedure, Statistical Analysis, and Anticipated Results. All of the authors contributed to frequent feedback sessions and the development of the project in general.

REFERENCES

- Adan, A. (1991). Influence of morningness–eveningness preference in the relationship between body temperature and performance: a diurnal study. *Pers. Individ. Dif.* 12, 1159–1169. doi: 10.1016/0191-8869(91)90080-U
- Adan, A., and Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: a reduced scale. *Pers. Individ. Dif.* 12, 241–253. doi: 10.1016/0191-8869(91)90110-W
- Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. *Chronobiol. Int.* 29, 1153–1175. doi: 10.3109/0742052.2012.719971
- Adan, A., and Guàrdia, J. (1993). Circadian variations of self-reported activation: a multidimensional approach. *Chronobiologia* 20, 233–244.
- Ansorge, C. J., and Scheer, J. K. (1988). International bias detected in judging gymnastic competition at the 1984 Olympic Games. *Res. Q. Exerc. Sport* 59, 103–107. doi: 10.1080/02701367.1988.10605486
- Antonakis, J., and Bastardo, N. (2013). *Swain: Stata Module to Correct the SEM Chi-Square Overidentification Test in Small Sample Sizes or Complex Models*. Chestnut Hill, MA: Boston College Department of Economics.
- Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010). On making causal claims: a review and recommendations. *Leadersh. Q.* 21, 1086–1120. doi: 10.1016/j.leaqua.2010.10.010
- Bai, L., Lao, S., Zhang, W., Jones, G. J., and Smeaton, A. F. (2007). “A semantic event detection approach for soccer video based on perception concepts and finite state machines,” in *Proceedings of the Eight International Workshop on Image Analysis for Multimedia Interactive Services* (Washington, DC: IEEE Computer Society), 30. doi: 10.1109/wiamis.2007.12
- Baltagi, B. H. (2011). *Econometrics*. New York, NY: Springer. doi: 10.1007/978-3-642-20059-5
- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73. doi: 10.1146/annurev.psych.57.102904.190152
- Browne, M. W., and Cudeck, R. (1993). “Alternative ways of assessing model fit,” in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Long (Beverly Hills, CA: Sage Publications), 136–162.
- Caci, H., Nadalet, L., Staccini, P., Myquel, M., and Boyer, P. (2000). The composite scale of morningness: further psychometric properties and temporal stability. *Eur. Psychiatry* 15, 278–281. doi: 10.1016/S0924-9338(00)00231-5
- Cartheeuw, P., Helsen, W., Gilis, B., and Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing. *J. Sports Sci.* 27, 1125–1136. doi: 10.1080/02640410903079179
- Cavanagh, P., and Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends Cogn. Sci.* 9, 349–354. doi: 10.1016/j.tics.2005.05.009
- Clarisse, R., Le Floch, N., Kindelberger, C., and Feunteun, P. (2010). Daily rhythmicity of attention in morning- vs. evening-type adolescents at boarding school under different psychosociological testing conditions. *Chronobiol. Int.* 27, 826–841. doi: 10.3109/07420521003794051
- Correa, Á., Molina, E., and Sanabria, D. (2014). Effects of chronotype and time of day on the vigilance decrement during simulated driving. *Accid. Anal. Prev.* 67, 113–118. doi: 10.1016/j.aap.2014.02.020
- Dosseville, F., Laborde, S., and Garnaczyk, C. (2014). “Current research in sports officiating and decision-making,” in *Contemporary Topics and Trends in the Psychology of Sports*, ed. C. Mohiyeddini (New York, NY: Nova Publishers), 13–38.
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fédération Internationale de Gymnastique [FIG] (2013). *Code of Points Rhythmic Gymnastics 2013–2016*. Lausanne: Fédération Internationale de Gymnastique.
- Fernandez-Villarino, M. A., Bobo-Arce, M., and Sierra-Palmeiro, E. (2013). Practical skills of rhythmic gymnastics judges. *J. Hum. Kinet.* 39, 243–249. doi: 10.2478/hukin-2013-0087
- Flessas, K., Mylonas, D., Panagiotaropoulou, G., Tsopani, D., Korda, A., Siettos, C., et al. (2015). Judging the judges' performance in rhythmic gymnastics. *Med. Sci. Sports Exerc.* 47, 640–648. doi: 10.1249/MSS.00000000000000425
- Gabrielsson, A., and Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007
- Glisky, E. L. (2007). “Changes in cognitive function in human aging,” in *Brain Aging: Models, Methods, and Mechanisms: Frontiers in Neuroscience*, ed. D. R. Riddle (Boca Raton, FL: CRC Press), 3–20.
- Gomes, A. A., Tavares, J., and deAzevedo, M. H. P. (2011). Sleep and academic performance in undergraduates: a multi-measure, multi-predictor approach. *Chronobiol. Int.* 28, 786–801. doi: 10.3109/0742052.2011.606518

- Greenwood, K. M. (1994). Long-term stability and psychometric properties of the Composite Scale of Morningness. *Ergonomics* 37, 377–383. doi: 10.1080/00140139408963653
- Hagemann, N., Strauss, B., and Leifßing, J. (2008). When the referee sees red. *Psychol. Sci.* 19, 769–771. doi: 10.1111/j.1467-9280.2008.02155.x
- James, L. R., and Singh, B. K. (1978). An introduction to the logic, assumptions, and basic analytic procedures of two-stage least squares. *Psychol. Bull.* 85, 1104–1122. doi: 10.1037/0033-2909.85.5.1104
- Lazarus, R. S. (2000). How emotions influence performance in competitive sport. *Sport Psychol.* 14, 229–252.
- Linacre, J. M. (1999). Estimation methods for Rasch measures. *J. Outcome Meas.* 3, 382–405.
- Looney, M. A. (2003). Evaluating judge performance in sport. *J. Appl. Meas.* 5, 31–47.
- Loula, F., Prasad, S., Harber, K., and Shiffrar, M. (2005). Recognizing people from their movement. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 210–220. doi: 10.1037/0096-1523.31.1.210
- Nevill, A. M., Balmer, N. J., and Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychol. Sport Exerc.* 3, 261–272. doi: 10.1016/S1469-0292(01)00033-4
- Oksama, L., and Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Vis. Cogn.* 11, 631–671. doi: 10.1080/13506280344000473
- Oudejans, R. R., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., and Beek, P. J. (2000). Errors in judging ‘offside’ in football. *Nature* 404, 33. doi: 10.1038/35003639
- Plessner, H. (1999). Expectation biases in gymnastics judging. *J. Sport Exerc. Psychol.* 21, 131–144.
- Plessner, H., and Schallies, E. (2005). Judging the cross on rings: a matter of achieving shape constancy. *Appl. Cogn. Psychol.* 19, 1145–1156. doi: 10.1002/acp.1136
- Roca, J. (2012). *Measuring the Three Attentional Networks in Vigilance Context and their Relationship with Driving Behavior*. Doctoral dissertation, University of Granada, Granada.
- Roca, J., Castro, C., López-Ramón, M. F., and Lupiáñez, J. (2011). Measuring vigilance while assessing the functioning of the three attentional networks: the ANTI-Vigilance task. *J. Neurosci. Methods* 298, 312–324. doi: 10.1016/j.jneumeth.2011.04.014
- Rodenberg, R. (2011). Perception? Reality: analyzing specific allegations of NBA referee bias. *J. Quant. Anal. Sports* 7, 8. doi: 10.2202/1559-0410.1326
- Schlegel, K., Grandjean, D., and Scherer, K. R. (2014). Introducing the Geneva emotion recognition test: an example of Rasch-based test development. *Psychol. Assess.* 26, 666–672. doi: 10.1037/a0035246
- Schlegel, K., and Scherer, K. R. (2015). Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): psychometric properties and construct validation. *Behav. Res. Methods* 47, 1–10.
- Schmidt, C., Peigneux, P., Leclercq, Y., Sterpenich, V., Vandewalle, G., Phillips, C., et al. (2012). Circadian preference modulates the neural substrate of conflict processing across the day. *PLoS ONE* 7:e29658. doi: 10.1371/journal.pone.0029658
- Smith, C. S., Reilly, C., and Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *J. Appl. Psychol.* 74, 728–738. doi: 10.1037/0021-9010.74.5.728
- Ste-Marie, D. M. (1999). Expert-novice differences in gymnastic judging: an information-processing perspective. *Appl. Cogn. Psychol.* 13, 269–281. doi: 10.1002/(SICI)1099-0720(199906)13:3<269::AID-ACP567>3.0.CO;2-Y
- Ste-Marie, D. M. (2000). Expertise in women’s gymnastic judging: an observational approach. *Percept. Mot. Skills* 90, 543–546. doi: 10.2466/pms.2000.90.2.543
- Steiger, J. H., and Lind, J. (1980). Statistically-based tests for the number of common factors. *Paper Presented at the Annual Spring Meeting of the Psychometric Society*, Iowa City, IA.
- Vast, R. L., Young, R. L., and Thomas, P. R. (2010). Emotions in sport: perceived effects on attention, concentration, and performance. *Aust. Psychol.* 45, 132–140. doi: 10.1080/00050060903261538
- Wright, B. D., and Mok, M. (2000). Rasch models overview. *J. Appl. Meas.* 1, 83–106.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 van Bokhorst, Knapová, Majoranc, Szebeni, Táborský, Tomić and Cañadas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.