A new statistical engine that can systematically pinpoint how accurately gymnastics judges apply required marking guidelines has been developed by researchers in Switzerland. The results are published in De Gruyter's *Journal of Quantitative Analysis in Sports*.

The results of gymnastics competitions are decided by marks assigned by expert judges and it has long been known that these judges do not always score competitors fairly and accurately. In their new study, Dr. Hugues Mercier from the Haute-École Arc Ingénierie and Sandro Heiniger from the University of St. Gallen used data from 21 international and continental gymnastics competitions held between 2013 and 2016, including the 2016 Rio Olympic Games, to investigate the accuracy and fairness of gymnastics judges.

Judging inaccuracies occur because a gymnastics routine includes a series of complex movements, and it is very hard for judges to evaluate every single element accurately. The marks are thus prone to inconsistencies. Meanwhile, fairness relates to impartiality and lack of favoritism. For example, it is well-established that sports judges, consciously or not, often display national bias, which manifests itself in a tendency to overmark competitors from their own country or undermark competitors from other countries.

The researchers first studied and quantified the intrinsic judging challenges of each apparatus and discipline, and how judges behave for each of them. They found that the pommel horse, for instance, is significantly harder to judge than the floor exercise. They also found that judges are more precise at judging the best athletes than mediocre ones. "This overall understanding has allowed us to evaluate the accuracy of judges compared to their peers, which had never been done before," Dr. Mercier said.

The most important conclusion of the study is that some judges are noticeably better at judging than others. "The vast majority of judges are fair and unbiased. However, judging is very hard, and even among the best-trained judges at the international level, there are judges who are, demonstrably and systematically, two to three times more accurate than their peers," Dr. Mercier said.

Following the study's analysis of judging behavior in gymnastics, the Fédération Internationale de Gymnastique (FIG), the global governing body for gymnastics, has already implemented systemic changes to improve the quality of judging, and mitigate the impact of judging errors. "By telling judges what they do well, and where they can improve, we are convinced that our work will lead to even better judging in the future," Dr. Mercier explained.
