

**Report of the FEI Dressage Judging Working Group**

Lausanne, 8 November 2018

## **TABLE OF CONTENTS**

<b>INTRODUCTION .....</b>	<b>2</b>
CHARGE GIVEN BY THE FEI BUREAU TO THE DRESSAGE JUDGING WORKING GROUP .....	2
DJWG METHOD OF OPERATION .....	2
COMPOSITION OF THE DJWG .....	3
ACKNOWLEDGEMENTS .....	4
<b>CHAPTER 1. BASIS FOR ALL RECOMMENDATIONS .....</b>	<b>5</b>
<b>CHAPTER 2. REVIEW AND ANALYSIS OF DRESSAGE JUDGING.....</b>	<b>6</b>
2.1 EXISTING FEI DRESSAGE JUDGING SYSTEMS .....	6
2.2 COMPARISONS TO OTHER DISCIPLINES & SPORT .....	7
2.3 JUDGING SYSTEMS USED IN REINING AND VAULTING .....	7
2.4 DIVISION OF JUDGING TASKS AND THE ROLE OF JUDGES .....	7
2.5 USE OF NEW TECHNOLOGY .....	10
2.6 COGNITIVE DEMANDS IMPOSED ON JUDGES .....	10
2.7 REVIEW OF OTHER SUPPORTING RESEARCH .....	13
2.8 TESTING A CODE OF POINTS SYSTEM: COPPSAG .....	14
2.9 ESTABLISHING A CODE OF POINTS: COP GROUP.....	15
2.10 REVIEW OF THE FEI DRESSAGE TASK FORCE 2009 RECOMMENDATION ON NUMBER OF JUDGES .....	16
2.11 SCORE CORRECTION MODELS .....	16
2.12 COLLECTIVE (SUMMARY) MARKS .....	17
2.13 REVIEW OF THE JUDGES EDUCATION SYSTEM .....	19
2.14 JUDGES DASHBOARD .....	20
2.15 REVIEW OF THE REPORT OF THE FEI DRESSAGE TASK FORCE 2009 .....	21
2.16 A CONTINUING PROCESS OF REVIEW AND REVISION .....	21
<b>ANNEX 1. DJWG RECOMMENDATIONS .....</b>	<b>23</b>
<b>ANNEX 2. MEETINGS OF THE DJWG .....</b>	<b>25</b>

## INTRODUCTION

### Charge given by the FEI Bureau to the Dressage Judging Working Group

The Dressage Judging Working Group (DJWG) were tasked by the FEI Bureau to;

- Provide a general evaluation and review of the current dressage judging system and make recommendations for improvement
- Identify relevant factors and protocols practised in other subjectively judged sports, applicable to the dressage judging system, and make appropriate recommendations
- Identify and make recommendations for improvement of the current FEI dressage judging system
- Evaluate, and make recommendations for improvement for, the measures introduced after the FEI 2009 Dressage Task force including: the introduction of 7 judges for major events, the introduction of a Judges Supervisory Panel, the introduction of half and the judge's education system

### DJWG method of operation




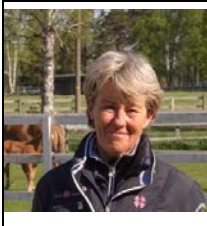
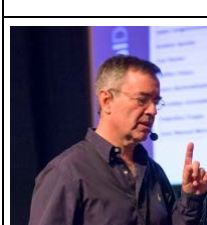
The DJWG commenced by acknowledging the brief provided by the FEI Bureau and noted the timeframe and intermediary deadlines for summary updates. They then addressed how they would operate and interact as a group in order to deliver the requirements set by the FEI. They decided to use a combination of physical meetings and conference calls and to develop a working document to record and update their findings.

In addition to this, some members of the DJWG may be delegated specific research or information-gathering projects which they would feedback to the group.

Where outside input was needed they would invite contributors to join conference calls and meetings. Working groups, acting as sub-groups, may be formed when necessary to provide information back to the DJWG from experts or where brainstorming required more lengthy input.

See [ANNEX 2](#) for the list of meetings of the DJWG.

## COMPOSITION OF THE DJWG

	<p>Chair  Frank Kemperman. NED  Member FEI Bureau  Chair of FEI Dressage Commission.  Organiser of the annual Aachen CHIO</p>
	<p>Bettina de Rham SUI  FEI Director of Dressage, Para Dressage, Reining and Vaulting</p>
	<p>Maribel Alonso MEX  FEI 5* Judge  Jury member OG London 2012  President of Jury OG Rio de Janeiro 2016  2013-17 Member of the FEI Dressage Commission</p>
	<p>Richard Davison GBR  Four time Olympian. European medallist. Fellow, International Dressage Trainer Club. Former Vice President, International Dressage Rider Club. Fellow British Horse Society. Former member FEI Dressage Committee. Visiting Fellow and Hon Doctorate of Science, Nottingham Trent University.</p>
	<p>Kyra Kyrklund FIN  Six time Olympian. World Championship individual medallist. President of the International Dressage Riders Club. Trainer of numerous national Teams. Fellow International Dressage Trainers Club. Former member of FEI Dressage Committee. Professor at the Swedish University of Agricultural Sciences</p>
	<p>Dr. David Stickland GBR  Princeton University Physicist at CERN  Founder of Global Dressage Analytics  Developer of the Judges Dashboard  Consultant to the FEI on Dressage Analysis</p>

## ACKNOWLEDGEMENTS

The DJWG would like to thank the following who have contributed towards the findings of this report:

- The team of the Dressage Department FEI HQ
- Anky van Grunsven, Raymond Grether, Erich Breiter, Stephen Clarke, Katrina Wüst, Dr Hughes Mercier
- Members of the NTU Group: Dr. Carol Hall, Dr. Alex Sumich, Dr. Christina Howard, Dr. Nadja Heym, Cassie White
- Members of the COPPSAG Group: Stephen Clarke (FEI Judge General), Katrina Wüst (5\* Judge/IDOC), Wayne Channon (IDRC), Uwe Mechlem (IDTC), Anky van Grunsven (Athlete) and Linda Keenan (IDTC)
- Members of the CoP Group: Stephen Clarke (Chair), Katrina Wüst, Hans-Christian Matthiesen, Maribel Alonso, David Hunt (President of the IDTC) and Kyra Kyrklund (President of the IDRC).

## CHAPTER 1. Basis for all recommendations

The DJWG acknowledged the FEI's duty to athletes in providing a fair judging system. They believed, in the context of awarding marks, fairness is expressed by consistency of judgement and decision making.

The DJWG have identified five fundamental principles which they consider should underpin a dressage judging system. All systems must satisfy these tests. All future proposals for improvement must also satisfy these five tests:

1. Be **TRANSPARENT**
2. Be **EASY TO UNDERSTAND** by all stakeholders
3. Be **FAIR** (applied consistently & objectively to all competitors)
4. Be **EASY TO USE** by all levels of judges (expert & non-expert) on a global scale
5. Be **EASY TO REVIEW AND ADJUST** (content of CoP and judging protocols)

The DJWG recommends that in order to deliver the duty of fairness to athletes all future proposals and initiatives should, as a minimum, satisfy all five of these criteria.

Currently, when making judging decisions (and awarding marks), judges do so either by:

- making their decisions individually, with no communication or collaboration with other officiating judges. This is the case for most FEI classes;
- making their decisions as members of a joint judging panel, when their decisions are joint and collaborative. This is commonly used for FEI competitions for young/age horse classes.

The DJWG focussed their investigations mainly into the scenario where judges are performing their roles individually. The arguments and conclusions below are in the context of a judge acting as a member of an officially appointed group of judges and thereby making judgements and decision alone, and not when acting as President of the Ground Jury where they may they have extra duties.

## CHAPTER 2. Review and analysis of dressage judging

### 2.1 Existing FEI dressage judging systems

In reviewing the current judging systems used by the FEI, aside from drawing on their own collective experiences and those from other external sources, the DJWG invited Katrina Wüst (KW) to deliver a review of certain aspects of judging. This review was based upon a presentation KW had made at the Global Dressage Forum Stakeholder meeting in Aachen July 2018 (in collaboration with Hans-Christian Matthiesen) and also at the International Dressage Trainer Seminar in October 2018. One of the main aspects KW focussed upon was the potential influence of bias and she lobbied for a clearer understanding of these issues between all parties directly affected by them, namely riders, trainers, organizers and the judges themselves.

#### Possible biases in judging

KW was of the opinion that judges are inevitably subject to possible biases in their work, judges need to be aware of them and attempt to limit their influence, but also judging systems should be configured to minimise the influence of such biases.

Example of such influences are:

- **Order bias:** A draw according to the World Ranking List raises expectations that the most successful (or better?) riders come at the end of the class. Judges have to realize that they might be biased by this starting order and remain fully aware of where the problems lie.
- **Conformity bias:** should not occur, however there are pressures, both real and imaginary, that can influence a judge to conform to what they think their colleagues may do.
- **Memory influenced bias:** should not happen, however, it is impossible for human beings not to memorize certain things, be it horses, riders, presentations or anything else. The judge has to try to 'switch off' his memory and to judge as impartially as possible.
- **Reputation bias:** Closely connected with conformity bias and starting order bias - the judge wants to stay 'in line' with their colleagues and gives good marks to famous riders who he thinks might deserve them.
- **Patriotism bias:** Is not acceptable, it has nothing to do with psychological phenomena that are common to all of us, but is a simple nationalistic attitude.
- **Anchoring effect:** is a cognitive bias for an individual to rely too heavily on an initial piece of information offered (known as the 'anchor') when making decisions.

In terms of regulating standards of judging, KW believed the judge's education system, together with the re-validation system for existing judges, should be more robust and effective.

The DJWG took note of this summary and acknowledged the need for the FEI to minimise the effects of any 'human' factors which could lead to an inappropriate slanting of marks.

## 2.2 Comparisons to other disciplines & sport

The DJWG compared the judging systems and processes used in other equestrian disciplines and also in some other non-equestrian sports. To this end the aim was to identify:

- Common processes between dressage and other sports/disciplines
- Common difficulties shared between dressage and other sports/disciplines
- Lessons to be learned how other sports/disciplines implement and address change
- Lessons to be learned how other sports/disciplines review their judging systems and procedures

## 2.3 Judging systems used in Reining and Vaulting

The DJWG met with experts from Reining and Vaulting. Anky van Grunsven and Raymond Grether who shared their knowledge of the judging system used in reining and Erich Breiter, who together with Bettina de Rham, the FEI Director of Vaulting, outlined the judging procedure used in Vaulting.

The main observations in both disciplines which were brought to the DJWG's attention were:

**Reining:** Scoring will be on a basis of 0-infinity, with 70 denoting an average performance. The individual manoeuvres are scored in ½ point increments from a low of -1 ½ to a high of +1 ½ with a score of 0 denoting a manoeuvre that is correct with no degree of difficulty. Competitions are judged with 1 to 5 Judges. At five-judges events the lowest and highest final scores are dropped. This high-low drop system is always used at Championships and Games.

**Vaulting:** further to a university research study conducted in 2009, which determined that each judge was not able to assess all the parameters they were tasked to judge (artistic score, technical score and horse score during a 1-minute performance), it was decided to divide the tasks of the judges in order for them to concentrate on one area at the time. Since the introduction of this new system in 2012, each judge is only tasked with evaluating either the technical score, the artistic score or the horse score. With this new system, the clear advantage is that each judge can concentrate on one aspect of each test and, in addition, this also gives a more accurate indication to the athletes and trainers of the areas they need to improve in their test.

## 2.4 Division of judging tasks and the role of judges

In their study of judging systems used in other sports the DJWG identified that division of judging tasks was sometimes used. This enabled judges to focus only on, and evaluate, specific aspects of the sporting performance.

As regards times when a judge's visibility is reduced and limited to only a part of the horse, while other parts are excluded, the DJWG considered whether the existing directives to judges need revision.

It is recognised that some other sports address lack of visibility by deploying the use of technology e.g. video replay or having linesmen. At FEI Championships video replay is used by the Judges Supervisory Panel, and this partially addresses the allocation of an inappropriate mark selected due to lack of visibility.



The DJWG consider it to be fundamental that each judge is only tasked with evaluating aspects for the horse/rider which they can actually see clearly. In studying the composition of existing dressage tests, together with their directives, such directives do not differentiate between as to what can actually be evaluated and what aspects cannot.

**RECOMMENDATION 1:** The DJWG recommends that all future standard judging directives and references (e.g. Code of Points etc) be formatted to ensure that they are relevant to the specific viewing position of each judge. The DJWG consider that judging directives and references should not be 'silent' on what is able to be observed/not observed from the various judging positions, as this leaves the priority of observations to the discretion of individual judges. Such discretion could lead to inconsistent judgements and a lack of fairness to athletes.

Generally, FEI dressage competitions focus on elite sport and, in that context, the DJWG defined that;

- the primary role of the dressage judge is to award the appropriate mark per movement
- the secondary role of the dressage judge is, as part of the OC/FEI Officials panel, to implement the FEI regulations

**RECOMMENDATION 2:** The DJWG recommends to the FEI that the judge's primary and secondary roles are more clearly defined and embedded in the job descriptions as recommended by the FEI Officials Working Group.

The DJWG considered any further roles of judges (examples being the education of the rider or to provide guidance on horse training) are subservient to the primary and secondary roles and are of lesser importance as far as the duties of FEI judges are concerned.

The DJWG considers that where FEI competitions are scheduled with the partial objective being the education of athletes, or the development of the horses, then the FEI should differentiate between these competitions compared to those aimed purely at the highest level of sport. The DJWG has not scrutinised judging systems used for athlete or horse development classes by the FEI but believe they might be better served by an adapted and specifically targeted judging and scoring system.

The DJWG observed that there may be a lack of clarity in the definition of competition levels and the purpose of different levels (Small Tour, Medium Tour, Big Tour) and of the different star levels. A discussion of these categories and levels should be instigated by the FEI.

**RECOMMENDATION 3:** The DJWG recommends to the FEI that, where appropriate, competition levels are redefined and their objectives more clearly identified.

In the context of deciding and awarding each mark, the DJWG concluded that judges do not act as part of a team, as is commonly suggested. In reaching this conclusion the DJWG rely upon the regulations prohibiting collaboration amongst judges when deciding upon and awarding each mark. The DJWG describes the work of a judge to be a non-collaborative and individual duty.

While a dressage test is in progress, the DJWG considers the fundamental duty of a judge commences and finishes at the commencement and completion of each movement as prescribed on the FEI dressage test sheet.

The DJWG noted that in terms of influencing the final ranking of each athlete, or that of the class as a whole, a judge's role is inherently limited by certain factors. Some examples of which are;

- the weighting of prescribed multiplying factors (coefficients) applied post-test by other officials;
- the influence of prescribed computer-generated scores e.g. Degree of Difficulty;
- the influence of third-party revision by panels such as the Judge's Supervisory Panel; decisions and recommendations by other Officials e.g. stewards, Appeal Panels

As the final ranking is outside of a judge's fundamental duty the DJWG are of the opinion that assessing how a judge has performed by relying solely on the final ranking of each judge is not an effective, nor accurate, method of evaluation. The DJWG recommends that when reviewing and assessing judging performance the focus is therefore placed on the relevance of each mark per movement.

This focus can be supported further by the format used to display and publish individual judge's marks. For example, consideration should be given to publishing the percentage variance of the overall score, or from judges who sit in a similar or close judging positions. The DJWG was of the opinion that the FEI should investigate the format of publication of class results so as to focus on (any) % score differences, instead of rank differences. In connection with this, the DJWG also recommend that the FEI place judging result columns based upon similar viewing positions (e.g. those judges positioned on the long side at E and B, and those judges positioned on the short side at H, M and C), so as to provide a more relevant comparison of marks allocated per judging/viewing position.

**RECOMMENDATION 4:** The DJWG recommends to the FEI that the focus on ranking should be reduced, especially for the purposes of evaluating judging performance. Instead the DJWG recommends that evaluation is assessed by scrutinising the appropriateness of each mark awarded for each movement.

**RECOMMENDATION 5:** The DJWG recommends to the FEI that the way results and marks are presented be amended to simplify comparison of results for judges officiating from similar viewing positions.

## 2.5 Use of new technology

The DJWG noted that the FEI currently employs a computer-based system to evaluate the degree of difficulty in some freestyle competitions. They concluded that, as technology develops and progresses, new innovations should be assessed to identify their use in dressage judging.

**RECOMMENDATION 6:** The DJWG recommended that technology within each judge's hut be established to facilitate the communication of observations regarding a limited number of breaches of regulation. These breaches should be strictly limited to prohibited equipment, the appearance of blood etc. Each judge should have the facility to alert the President of the Ground Jury that action needs to be taken during the test.

*The proposal was accepted at the FEI GA 2017 for implementation 2018 onwards.*

**RECOMMENDATION 7:** The DJWG recommends the FEI set up a sub-group to provide advice on development, innovation and use of modern technology in order to assist judges and improve accuracy and therefore fairness to athletes. The DJWG recommends that new and existing technology is reviewed to, among other things, assess the impact upon cognitive load for judges

The DJWG were made aware of the following innovations and recommends that the FEI investigate them as to their merits and suitability in the context of dressage judging:

- Evaluation of the music used by athletes in freestyles – various proposals have been received
- The use of sensors to measure certain movements.
- E-scribe where no paperwork is used by scribes. This is already tested at some events.

## 2.6 Cognitive demands imposed on judges

In considering the task imposed upon FEI judges, the DJWG felt they needed more expert advice regarding the cognitive demands that judges experience during competition. This included seeking evidence-based knowledge on human attention capability, concentration, memory, bias, perception, vision and other parameters.

Representing the DJWG, Richard Davison (RD) held a preliminary meeting with Nottingham Trent University (NTU) to explain the DJWG's requirements. Following this introductory meeting, a group of key researchers at NTU; with a wide range of expertise across many areas of cognitive psychology including memory, visual search, attention capacity and cognitive processing, was set up.

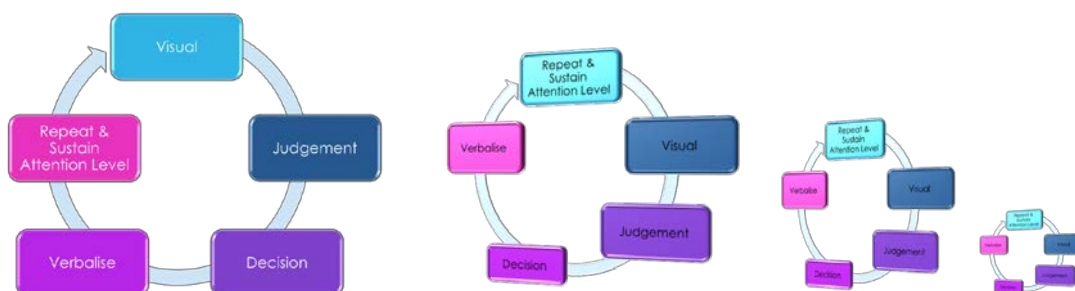
The NTU research group explored the human skills needed for optimal and consistent judge performance and, aside from the obvious issue of different viewing positions, they reviewed some likely factors which may lead to potential inconsistencies in delivery of judgements. Existing research in cognitive psychology, especially the studies based on the performance and judgements of sports officials, was reviewed and current literature relevant to dressage judging evaluated.

The NTU Group reviewed video footage showing a number of consecutive dressage movements from the FEI Grand Prix test. They concluded that repeated and sustained attentional level was required by dressage judges and that the number of decisions and

movements presented in a test created a complex and demanding cognitive processing load. They discussed the potential impact of the scheduling and length of FEI classes, including championship classes being held over two days.

The NTU Group summarised the existing general cognitive processing pathway of dressage judging as having four key stages;

- **Visual and auditory cognition** - where a judge scans and processes visual cues and uses auditory information from the horse, rider, arena, music, test papers, scribe and other parts of their environment.
- **Judgement** - where a judge utilises the information retained within their working memory, sourced from their visual and auditory search, and compares this to a reference embedded and retrieved from their long-term memory (e.g. FEI regulations / Judges Manual/previous images and previous personal knowledge and experience).
- **Decision making** - where the judge concludes their judgement and decides how to express it in terms of marks, penalties and/or actions.
- **Verbalisation & interaction** – where the judge verbalises their decision including the allocated marks to the scribe, often while also dictating a summary description of the observation and/or advice for the benefit and education of the rider.
- **Repeat and Sustain Attention Level** - whilst verbalising this information, the judge may also have to answer questions on clarification from the scribe or computer operator, whilst at the same time still continuing to observe, evaluate and memorise actions within the current and ongoing movements, or those movements already partially or fully completed.



The NTU research group examined the existing standard references for judges, namely the FEI Regulations for Dressage and The Judges Manual. These consist of detailed and lengthy written regulations and descriptions and the group felt that it was unlikely that judges could recall the full details from their memory and apply them consistently. It was therefore concluded that cognitive 'short-cuts' would be employed by judges, and that these could be based on a mixture of personal knowledge and experiences gained from events such as education courses, judging discussions, training, practical dressage riding, previous performances, spectating and numerous other sources. The group considered research showing that whilst drawing on previous experiences can improve the practical skill of a judge, without a standardised reference, it can also account for variability.

The current education process for judges includes the use of visual imagery in the form of 'live' dressage performances and also video footage. The NTU Group identified that the use of visual educational material could be highly effective especially in the context of dressage judging training.

The NTU group agreed that the DJWG's proposed development of a Code of Points reference would be an important and positive step, providing its format uses evidence-based principles to optimise each judgement. To this end the NTU group recommended the Code of Points should;

- be formatted to optimise perceptual cognition processes, including effective visual imagery
- define observable characteristics
- define how each characteristic is to be evaluated, quantified or measured
- define the visual scan required to observe each defining observable characteristic
- consider the use of visual education material as key part of the training of judges (based on the approach outlined above)

The NTU group also cited research evidencing factors which can negatively influence perceptual cognitive consistency. These included, among other things, the timing of the dressage class, the length, number and duration of breaks and the judge's physical and cognitive activity within the breaks. Within each dressage test there is the potential for the temporal position of movements to impact on how consistently these are assessed and for this to vary in relation to the factors noted above. Consideration should also be given to judging in different time zones, the potential impact of jet lag, sleep deficiency, diet and hydration. Existing medical conditions, the use of some medications, eyesight conditions and limitations and the use of stimulants may also have an effect on cognitive ability and consistency.

The group explained that the longer the duration of the class, especially when combined with the greater number of judgements per time period, then the greater the likelihood that some of these factors could impact on efficiency of decision making and judgement. RD explained that, in the context of dressage judging, the DJWG believed that consistency of judgements underpinned the obligation of fairness to athletes.

The NTU group cited evidence of other sports' governing bodies which have already started addressing perceptual cognitive efficiency of their officials, along with some professions, vocations and sectors within industry.

In summary, the NTU Group concluded that dressage judging involves a perceptual cognitive function and the fundamental requirements for delivering effective and consistent judging were:

- A standard reference (Code of Points) defining mark allocation and deductions formatted to optimise perceptual cognition function.
- Where possible a standardised visual search routine used to evaluate each definable observable characteristic for each movement.
- An evaluation of the perceptual cognitive load upon judges, especially when compiling and drafting dressage tests and scheduling classes and breaks and their duration in order to assist consistency of judgements.
- An objective and regular review of each judge's performance together with a relevant and targeted education system, including advice for judges on how to optimise their own cognitive performance.
- An objective and routine review of the Code of Points, judging processes and other judging protocols.

The NTU group acknowledge that the sport of dressage is more than just a sum of the parts and the scoring system needs to reflect this. The feedback from the research group should help to provide information towards an objective, sustainable and transparent coding system but this will require technical input from dressage experts.

The NTU Group's findings and recommendations were communicated to the DJWG by RD. They were also summarised and, together with those of the DJWG, presented by RD to representatives of the FEI and its Stakeholders at the FEI Sports Forum in April 2018.

The DJWG are indebted to and acknowledged the valuable input of the NTU Group and accepted their recommendations. They have also instructed the NTU Group to provide guidance during the compilation and format of the Code of Points and also to provide objective testing where appropriate.

The members of the NTU Research Group were;

Dr. Carol Hall, Associate Professor in Equine Behaviour & Welfare, School of Animal, Rural and Environmental Sciences

Dr. Alex Sumich, Associate Professor in Mental Health and Biopsychology, School of Social Sciences

Dr. Christina Howard, Associate Professor, School of Social Sciences

Dr. Nadja Heym, Psychology Dept, School of Social Sciences

Cassie White, Employability Manager, School of Animal, Rural and Environmental Sciences

Richard Davison, Olympic dressage rider and NTU Visiting Fellow

**RECOMMENDATION 8:** The DJWG recommends that the FEI publishes guidelines for judges on how to optimise their cognitive function and what factors can affect it both positively and negatively.

**RECOMMENDATION 9:** The DJWG recommends that the FEI provides evidence-based guidelines to Organising Committees regarding the duration of breaks and other appropriate requirements for judges to support and enhance optimal cognitive function

## 2.7 Review of other supporting research

Mercier & Klahn (2017) examined the performance of international gymnastic judges which provided useful comparative findings to dressage judging. The DJWG invited Prof. H. Mercier (HM) of the University of Neuchatel to present his findings.

Hugues Mercier and his colleague Sandro Heiniger made reference to their research paper : "Judging the Judges: A General Framework for Evaluating the Performance of International Sports Judges ". The abstract of which is partially reproduced here:

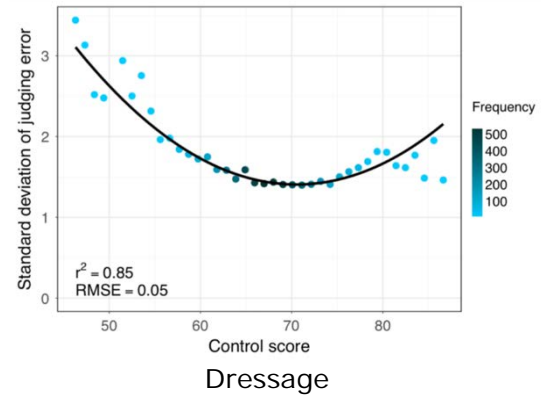
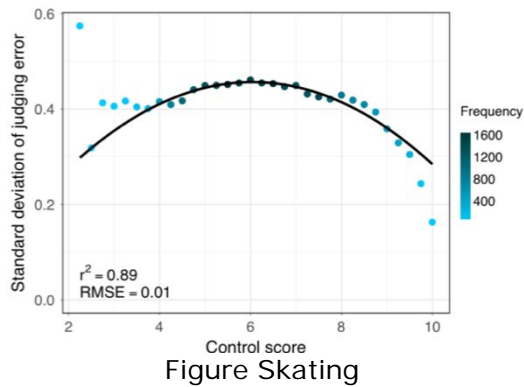
*Abstract—The monitoring of judges and referees in sports has become an important topic due to the increasing media exposure of international sporting events and the large monetary sums involved. In this article, we present a method to assess the accuracy of sports judges and estimate their bias. Our method is broadly applicable to all sports where panels of judges evaluate athletic performances on a finite scale.*

*We analyze judging scores from eight different sports with comparable judging systems: diving, dressage, figure skating, freestyle skiing (aerials), freestyle snowboard (halfpipe, slopestyle), gymnastics, ski jumping and synchronized swimming. With the notable exception of dressage, we identify, for each aforementioned sport, a general and accurate pattern of the intrinsic judging error as a function of the performance level of the athlete.*

...

*Our analysis also leads to valuable insights about the judging practices of the sports under consideration. In particular, it reveals a systemic judging problem in dressage, where judges disagree on what constitutes a good performance.*

Essentially for all other sports studied, the agreement between judges improves for higher scores and is worse for intermediate scores, but the behaviour is opposite for Dressage. As in the following example from Figure Skating (Left) and Dressage (Right).



This result agrees with the results of DS, also presented to the Working Group, that show the same behaviour for movement by movement standard deviation between judges. Best agreement for scores is in the range 6.5 to 7.5 and worse agreement both above and below that range. To quote from their conclusion:

*"Our analysis of dressage judges further shows that they increasingly disagree as the performance level increases, indicating a significant amount of subjectivity in the judging process compared to other sports with similar judging systems".*

Based upon this, and other, evidence it is the conclusion of the DJWG that the higher and lower end of the marks range lacks consensus

## 2.8 Testing a Code of Points system: COPPSAG

Based on their study of other judging systems and on the detailed reports from external experts such as NTU and Mercier, **the DJWG concluded that the practical implications of a Code of Points needed to be evaluated.** To this end they initiated a prototype study that was delegated to stakeholders with expert knowledge of dressage and judging.

They invited members of the Stakeholder Groups to recommend members with expert knowledge of dressage judging to join a working group which became known as the Code of Points Advisory Group (COPPSAG).

The COPPSAG members were Stephen Clarke (FEI Judge General), Katrina Wüst (5\* Judge/IDOC), Wayne Channon (IDRC), Uwe Mechlem (JSP), Anky van Grunsven (Athlete) and Linda Keenan (IDTC)

The DJWG briefed the COPPSAG to draft a sample Code of Points which;

- quantifies a series of deductions starting from 10 for a specified series of movements from a test prescribed by The DJWG (e.g. GP - from extended trot, passage, piaffe, passage, passage-piaffe-passage transitions, extended walk);
- identifies and includes only measurable observations which are easily visible to each judge, and to recommend wording (of the observations) that reflects everyday language and refers to specific parts of the horse/rider (to avoid ambiguity);
- prioritises scale of deductions which reflects faults ranging from serious to minor and which relate to the principle purpose of each movement;
- quantifies deductions which relate to the fault being observed either: momentarily, repetitively or continuously through each movement.

The FEI Judge General Stephen Clarke presented COPPSAG's findings to the DJWG at their in-person meeting in Amsterdam in January 2018. The DJWG accepted the draft format for the CoP and expressed their gratitude to COPPSAG for their expertise and hard work.

This led to the establishment of a CoP group tasked to deliver a full Code of Points for the grand prix movements.

## 2.9 Establishing a Code of Points: CoP Group.

In order to develop and expand the CoP further the DJWG delegated this to a sub-group of the Education Working Group. This group became known as the CoP Group and consisted of Stephen Clarke (Chair), Katrina Wüst, Hans-Christian Matthiesen, Maribel Alonso, David Hunt (President of the IDTC) and Kyra Kyrklund (President of the IDRC). Richard Davison assisted the Chair with administration. The NTU Group provided guidance and testing in order to ensure the format of the CoP optimized perceptual cognition.

The brief to the CoP Group is as listed below;

- Define a Code of points for each of the Grand Prix movements. The Code of Points should be:
  - flexible and easily adjustable
  - read as one complete template for each movement, comprising of the definition, the Code of Points, deductibles and firewalls
- Language used must be everyday English, easy to understand. The words used must relate to observations, and each observation has to be measurable in terms of scores
- The CoP has to be restricted to the priorities of the movements; requiring fewer comments from the judge and a max. four priorities per movement.
- The template for each movement should commence with a revised and modernised wording based on the current rules, which describes the perfect movement, followed by the description of each mark, to include half marks when applicable.
- In addition, the following should be defined:
  - Deductions should be faults not included in the criteria of each movement as described above and need to be listed, to complement the definitions.
  - In addition to the deductions, define a list of more or less serious faults (degree of fault and duration of the fault within the movement).
  - A component for the effect of the Rider must be included within each movement in the deductions (e.g. visibility of the aids).
- Firewalls to be included. A Firewall is a barrier score, in the presence of certain faults in a movement's presentation it will not be permissible to give a score higher than this firewall.
- Timeline: the new Code of Points should be tested in 2019, the Education system updated in accordance and the Code of Points Judging system should be introduced in 2021, following the 2020 Olympic Games.

The CoP Group commenced by each member being allocated specific movements from the current grand prix test. They were charged with drafting the main priorities of each movement as expressed by observables and then assigning marks out of ten to each variation of the prioritised observations.

At the date of publication of this report the CoP Group are continuing to progress the drafting of the Code and testing is planned during 2019.



**RECOMMENDATION 10:** The DJWG recommends to the FEI that they proceed with the draft Code of Points and, subject to testing and review, roll it out for use in all relevant classes. The DJWG recommend that the proposed Code of Points is supported, where possible, by appropriate visual imagery.

**RECOMMENDATION 11:** The DJWG recommends that the Code of Points should be formally reviewed by a panel of dressage experts every four years and that this review becomes a regulatory obligation and be budgeted for accordingly.

## 2.10 Review of the FEI Dressage Task Force 2009 recommendation on number of judges

The DJWG studied the impact of various judging panel compositions, and specifically how the number of judges impacted on consistency and fairness to athletes using analysis of results from previous Olympic Games and World Equestrian Games.

The utility of 7 judges has been demonstrated in analysis at Olympic Games. As the number of judges is reduced then, by definition, the relative importance and weight of each judge in the final ranking becomes more important. The effect on final ranking of removing one judge at a time from seven judges is considerably less important than that of removing one judge from five. The non-contentious nature of the final results of championships since the introduction of the 7-judge requirement, strongly supports the continuation of this rule. Such examples clearly show that the FEI should make it a goal to have as many judges (max 7) present in any given competition as is economically practical.

**RECOMMENDATION 12:** The DJWG recommends to the FEI that the current practice of using 7 judges at Senior Championships and Games, as defined in the FEI Dressage Rules, is continued.

## 2.11 Score correction models

Aside from the powers given to the Judges Supervisory Panel, the DJWG wanted to investigate the merits of a number of options aimed at mitigating and correcting a lack of consensus among judges. In summary these included studying two models;

### Percentage threshold correction model

The aim of this model is to adjust the highest and/or lowest outlying scores by comparing them against the next closest mark, or the average mark, awarded by other judges. The model relies upon a prescribed % threshold (e.g. 6%). Having received proposals based upon different %s from Stakeholder Clubs, the DJWG studied statistics provided by DS using historical data to assess the impact of a variety thresholds. They also took note of the model used by the Italian NF which is based upon a 4% threshold

As well as concluding that this correction method is not appropriate for judging panels of only three judges, the DJWG find that these models do not meet their test of fairness to all athletes within the same class. Whichever threshold figure is chosen those athletes whose score falls just short of the threshold are excluded from benefiting from the correction protocol and as such their overall result is excluded from the benefits of this provision. The DJWG are of the view that, in order to satisfy the principle of fairness, such

correction protocols must not be arbitrary but instead they must be applicable to all athletes within the same class.

While considering this model the DJWG also evaluated Article 438 and the impact of the current percentage threshold of 6% as used by the JSP at championships.

**RECOMMENDATION 13:** The DJWG recommends to the FEI that score correction models that are based upon a percentage threshold should not be used as they fail the DJWG's test of equal treatment and fairness to all athletes within the same class.

### Hi/Lo score removal model

The DJWG also studied what has become known as the Hi/Lo proposal. This is where the highest and lowest scores of the whole judging panel are removed. The DJWG found that this did meet their test of fairness as it was applied to all athletes within the same class. They believed that for the purpose of mitigating the impact of outlying scores, when used on an individual movement basis this could provide a short-term protection to athletes.

The DJWG concluded that until a Code of Points was available to assist in improving consensus, accuracy and consistency of mark per movement selection, then a Hi/Lo per movement was a fair protection model for athletes. However, this proposal did not gain support from some NFs and therefore the proposal was not presented at the FEI GA 2017

**RECOMMENDATION 14:** The DJWG are of the opinion that if a score correction model is to be employed then the model based upon Hi/Lo drop per movement provides the highest level of fairness as it applies to all athletes within the same class.

*However, the DJWG recommends that the FEI wait until the impact of the proposed Code of Points is evaluated before assessing the need for score correction models.*

## 2.12 Collective (summary) marks

Collective marks are awarded after the completion of the test and are a summary of specified aspects divided into categories.

Taking into account the information provided on reliable cognitive function, together with the cumulative impact of the duration and other challenges of dressage competitions, the DJWG identified the implementation and use of collective marks as an area of potential inconsistency.

They scrutinised various data models which both included and excluded collective marks and found that for most athletes it made no significant difference to their final ranking whether they were included or not. However, there was evidence of some significant differences in ranking to some athletes. The DJWG is of the view that the same degree of fairness is owed to all athletes, especially those within the same class.

A greater degree of accuracy and precision of mark selection is more likely to occur if the mark is selected as soon as possible after the completion of each movement. Mark selection decided retrospectively requires a greater cognitive load in memory retrieval and has more potential to be influenced by other cognitive influences such as types of bias, especially when the cumulative impact of the duration of dressage competitions is taken into account.

On the basis of providing greater potential fairness to athletes and also optimising consistent cognitive output of judges, the DJWG therefore recommended to the FEI that three out of the four collective marks should be removed.

In future judging systems should either be based entirely on summary marks (awarded post-test) or at the completion of each movement. The latter is the DJWG preferred system.

The justification for the rider mark remaining is that currently the facility to evaluate the components of this mark are not provided during the course of the test. However, any new Code of Points should include this evaluation throughout the course of the test.

Aside from providing greater potential of fairness to athletes, the removal of collective marks provides benefit to the public, the judges and organising committees.

The DJWG proposal to remove three collective marks was approved at the General Assembly 2017 and introduced on January 1 2018.

Statistical analysis of the impact of removal, after first 6 months in 2018, compared to 2017, showed that there is neither a statistically significant difference in average scores assigned nor in the boost given by the collective marks compared to the technical marks. In 2017 the average Collective score was  $69.2 \pm 0.1$  while in 2018 it was slightly higher at 71.8. Average total scores in 2017 were  $67.5 \pm 0.1$  and in 2018 they were  $67.3 \pm 0.1$ . For all the reasons outlined above the DJWG remains convinced that the change made for 2018 was justified and should be continued.

**RECOMMENDATION 15:** The DJWG recommends to the FEI that summary marks awarded post-completion of the test be discontinued. Their reasoning for this is that, as this relies upon a longer-term memory retrieval when compared to awarding a mark at the conclusion of each individual movement, there is a lower degree of probability of precision when selecting the appropriate mark and leaves too much to the discretion of each individual judge. When considering the length of many FEI classes the impact of the cumulative cognitive load reduces the probability of accurate mark selection and increases the probability of inconsistency thus increasing the likelihood of unfairness to some athletes.

*The FEI accepted the recommendation of the DJWG to remove three out of the four standard 'collective' marks at the FEI GA 2017.*

## 2.13 Review of the Judges Education System

The aim of the FEI Judges Education system is to provide global standards for education and certification of FEI judges. Qualification courses and refresher seminars on a face-to-face or online basis will provide opportunities for promotion and status maintenance. It is important to have continued exchange between the FEI, the NFs and the judges with regards to the regional situation in order to establish a rolling education calendar and proactively address educational developments.

To date, the Job Descriptions and Checklists for Judges (and all other FEI Officials) have been drafted and submitted to the Technical Committee for feedback/approval and are currently being reviewed by the Education and Legal Departments. Once they have been approved, these can be integrated into the Education System.

All current course material has been categorized and reviewed for its usability. The selected course material will be formatted with current FEI branding. The finalised course content/presentations will be submitted to the Technical Committees for review and approval with the aim of implementation for the first courses in 2019.

An online Foundation Course for Officials is currently being developed for all disciplines with the aim of providing a solid "foundation" for all FEI Officials in the areas of; History of the FEI and the Disciplines, FEI Officials, Departments and Committees, General Rules and Regulations, Code of Conduct for Horse Welfare, Environment and Officials, the FEI Legal System, Statutes, Values and Horsemanship.

Once the Job Descriptions and Checklists have been finalised, these will be integrated into the Education System. Following on from this will be the production and/or update of course syllabi (on-site) based on Job Descriptions for Judges with some online elements with the objective of producing standardised course material for Judges Education including Exams. The appointment of a Course Director General (CDG) for Judges in each discipline is currently in progress to represent both the Course Directors and the Pedagogical Coach (to be appointed to train CDGs).

The Education department has allocated a budget for updating the Education System. As the Education System is drawn up, contracts will be made with those who will produce the course structure, content and exams. This will be used in particular for the Online Foundation Course and any other updates required to onsite/online Course Content with the aim of implementing this for the first courses in 2019.

The DJWG are of the opinion that the Code of Points should become central to the education programme. They believe the Code of Points will provide a more focussed education plan and achieve a more effective outcome. They also support the FEI Officials Working Group's recommendation that increasing the educational skill of course educators is fundamental and resources to validate educators should be made available by the FEI

**RECOMMENDATION 16:** The DJWG recommends to the FEI that the Education Working Group be tasked to embed the proposed Code of Points into the education system along with those relevant recommendations from the FEI Officials Working Group report

Any standard judging reference must be relevant to the viewing position to the judge. In other words it must not apply to characteristics they cannot actually observe.

The DJWG also recommends to the FEI that part of the education programme should be dedicated to maximising the performance potential of each judge. This requires acknowledging the various factors that can increase or impair cognitive function and which all judges should be obliged to address in order to deliver their role optimally and also their duty of fairness to athletes. Declaration of medical conditions, inadequate eyesight, use of medications etc, and other personal factors which might affect concentration and attention, should be obligatory.

**RECOMMENDATION 17:** The DJWG recommends to the FEI that guidelines should be provided to judges to make them aware of factors that can impact upon their cognitive performance when judging, such as attention and concentration.

The DJWG recommends that, in the interest of fairness to the Athlete, the FEI seeks further advice on the impact of medical conditions and medications upon judging decisions.

## 2.14 Judges Dashboard

The Judges Dashboard is a web interface that judges and supervisors can use to obtain both an overview of judging, but also drill down to specifically study individual competitions, rides and movements. Its purpose is twofold;

- Firstly, to give the FEI and its supervisory panels such as the JSP and overview of each judges work history. Analytics such as mean difference, standard-deviation of difference and number of differences larger than 5% can be calculated for individual competitions, date ranges, or indeed specifically compared with other named judges. This type of data can be useful in reviewing promotions, championship judges, education etc.
- Secondly, the judges themselves can access all data related to their own analytics. They can see how they compare with their peers, they can also tunnel into specifics of individual rides and movements to understand better where any differences in score may have arisen.

It is intended as both an educational and a supervisory tool.

The DJWG studied the Judges Dashboard which was fully deployed in mid 2017. The DJWG were informed that this includes all movement and judge scores for all FEI levels as from January 1 2017, and all grand prix level scores since 2008. All FEI judges have received accounts whereby they can access their statistics and delve into details of individual competitions or tests.

Since January 2018 judges receive email notifications each time there are new results for them, only one judge has opted out of receiving these notifications. Many judges have responded favourably to the emails and dashboard information. Detailed login tracking is not yet deployed.

Computer systems have been upgraded to improve response times and since April 2018 it has been possible to access their dashboards from an FEI Login, thus removing the need for separate account names and passwords.

The FEI judge supervisor dashboard is now being made available, also via the FEI login, to the members of the JSP. Video training sessions have been established. JSP members will also be invited to such sessions.

Uploading of results on a weekly basis is efficiently managed by the FEI dressage office.

A first evaluation is planned for 2019

**RECOMMENDATION 18:** The DJWG recommends to the FEI that they review the utility of the Judge's Dashboard in 2019 and how it can be used to support the education of judges

## 2.15 Review of the report of The FEI Dressage Task Force 2009

The DJWG reviewed the relevant section of the report of The Dressage Task Force 2009. The section which was considered relevant for the DJWG was section 6 entitled Judging. Many of the proposals had already been adopted by the FEI and have therefore been dealt with above. They include the education and evaluation of judges, the creation and role of the JSP, the division of tasks, the number and position of judges and the introduction of half marks.

As for half marks, these have been widely adopted and are firmly established in international and many national dressage competition levels. It is the hope of the DJWG that the Code of Points currently in preparation will formalise the use of half marks to give even more precision to the movement scores.

The DJWG were of the opinion that the implemented recommendations, relevant to dressage judging, of the FEI Dressage Task Force 2009 had a positive impact upon judging. They consider that, among other things, a Code of Points as recommended within this report, will reinforce and provide greater cohesion to the 2009 recommendations.

## 2.16 A Continuing process of review and revision

The DJWG considered the role of the Judges Supervisory Panel (JSP) and of the FEI Judge General. They were of the opinion that a continued process of review and revision is fundamental and recommend that the FEI establish a permanent Judging Advisory Group which combines and expands these two roles

A Judging Advisory Group (JAG) should be formed and charged to;

- Include the routine review of the Code of Points
- Supervise, monitor and review the performance of all judges by means of objective data such as the Judges Dashboard system
- Recommend to the FEI DC judges and the Judging Supervisory Panel members for championships and other high-profile events based on evidence of current performance data and objective selection.
- Recommend the appointment of Educators of Judges and standardise delivery of judge education and content.
- Advise on all judging relating issues
- As to the composition of JAG the DJWG recommend;
  - It is comprised of experts proposed by the Stakeholder Clubs
  - It should have a minimum of three members and a maximum of five members of whom a minimum of two and a maximum of three should be FEI 5\* judges (active or retired). Trainers and riders who have previously performed at an Olympic or World Championships and are still active within the sport should be included

- The term of appointment would be for four years with a maximum of three terms
  - There is no upper age limit
- It should be chaired by the FEI Judge General whose title should be changed to Chairman of JAG
- In the case that the Chairman of JAG is not a member of the FEI Dressage Committee, then he/she should be invited to attend meetings of the FEI DC
- Potential conflicts of interest, especially self-interest, among members of JAG should be made transparent and managed by protocol.
- New job descriptions should be created for the JAG Chair and members

**RECOMMENDATION 19:** The DJWG recommends to the FEI that a Judging Advisory Group be established as an evolution of the current Judging Supervisory Panel and incorporating also the role currently held by the FEI Judge General

## ANNEX 1. DJWG RECOMMENDATIONS

1. The DJWG recommends that all future standard judging directives and references (e.g. Code of Points etc) be formatted to ensure that they are relevant to the specific viewing position of each judge. The DJWG consider that judging directives and references should not be 'silent' on what is able to be observed/not observed from the various judging positions, as this leaves the priority of observations to the discretion of individual judges. Such discretion could lead to inconsistent judgements and a lack of fairness to athletes.
2. The DJWG recommends to the FEI that the judge's primary and secondary roles are more clearly defined and embedded in the job descriptions as recommended by the FEI Officials Working Group.
3. The DJWG recommends to the FEI that, where appropriate, competition levels are redefined and their objectives more clearly identified.
4. The DJWG recommends to the FEI that the focus on ranking should be reduced, especially for the purposes of evaluating judging performance. Instead the DJWG recommends that evaluation is assessed by scrutinising the appropriateness of each mark awarded for each movement.
5. The DJWG recommends to the FEI that the way results and marks are presented be amended to simplify comparison of results for judges officiating from similar viewing positions.
6. The DJWG recommended that technology within each judge's hut be established to facilitate the communication of observations regarding a limited number of breaches of regulation. These breaches should be strictly limited to prohibited equipment, the appearance of blood etc. Each judge should have the facility to alert the President of the Ground Jury that action needs to be taken during the test.
7. The DJWG recommends the FEI set up a sub-group to provide advice on development, innovation and use of modern technology in order to assist judges and improve accuracy and therefore fairness to athletes. The DJWG recommends that new and existing technology is reviewed to, among other things, assess the impact upon cognitive load for judges.
8. The DJWG recommends that the FEI publishes guidelines for judges on how to optimise their cognitive function and what factors can affect it both positively and negatively.
9. The DJWG recommends that the FEI provides evidence-based guidelines to Organising Committees regarding the duration of breaks and other appropriate requirements for judges to support and enhance optimal cognitive function
10. The DJWG recommends to the FEI that they proceed with the draft Code of Points and, subject to testing and review, roll it out for use in all relevant classes. The DJWG recommend that the proposed Code of Points is supported, where possible, by appropriate visual imagery.



11. The DJWG recommends that the Code of Points should be formally reviewed by a panel of dressage experts every four years and that this review becomes a regulatory obligation and be budgeted for accordingly.
12. The DJWG recommends to the FEI that the practice of using 7 judges at Championships and Games is maintained.
13. The DJWG recommends to the FEI that score correction models that are based upon a percentage threshold should not be used as they fail the DJWG's test of equal treatment and fairness to all athletes within the same class.
14. The DJWG are of the opinion that if a score correction model is to be employed then the model based upon Hi/Lo drop per movement provides the highest level of fairness as it applies to all athletes within the same class. However, the DJWG recommends that the FEI wait until the impact of the proposed Code of Points is evaluated before assessing the need for score correction models.
15. The DJWG recommends to the FEI that summary marks awarded post-completion of the test be discontinued. Their reasoning for this is that, as this relies upon a longer-term memory retrieval when compared to awarding a mark at the conclusion of each individual movement, there is a lower degree of probability of precision when selecting the appropriate mark and leaves too much to the discretion of each individual judge. When considering the length of many FEI classes the impact of the cumulative cognitive load reduces the probability of accurate mark selection and increases the probability of inconsistency thus increasing the likelihood of unfairness to some athletes.
16. The DJWG recommends to the FEI that the Education Working Group be tasked to embed the proposed Code of Points into the education system along with those relevant recommendations from the FEI Officials Working Group report.
17. The DJWG recommends to the FEI that guidelines should be provided to judges to make them aware of factors that can impact upon their cognitive performance when judging, such as attention and concentration. The DJWG recommends that, in the interest of fairness to the Athlete, the FEI seeks further advice on the impact of medical conditions and medications upon judging decisions.
18. The DJWG recommends to the FEI that they review the utility of the Judge's Dashboard in 2019 and how it can be used to support the education of judges .
19. The DJWG recommends to the FEI that a Judging Advisory Group be established as an evolution of the current Judging Supervisory Panel and incorporating also the role currently held by the FEI Judge General

## ANNEX 2. Meetings of the DJWG

- 2016 conference calls:
  - 25.7.16,
  - 22.8.16
  - 28.11.16
  
- 2017 in person meetings:
  - 26-27.01.17
  - 22.05.17
  
- 2017 conference calls:
  - 21.02.17,
  - 01.03.17,
  - 23.03.17,
  - 20.04.17,
  - 30.05.17,
  - 19.06.17,
  - 27.06.17,
  - 31.07.17,
  - 20.09.17,
  - 09.10.17,
  - 06.12.17
  
- 2018 in person meetings:
  - 16.01.18.,
  - 12-13.04.18
  
- 2018 conference calls:
  - 19.03.18,
  - 23.05.18,
  - 26.06.18,
  - 10.07.18,
  - 08.10.18,
  - 05.11.18