

Presentation of the *Scoresheet Proposal'17*

Despite the earnest efforts of all the jurors, there is a feeling that something could be done to improve the scoring system of IYPT. There are a lot of factors which influence the quality of grading at IYPT, like:

- Common conceptual framework about what are the tasks for *Rep, Opp, Rev*,
- Tentative consensus among the jurors about what is to be evaluated and what should have greater weight in the final grade,
- Tentative consensus among the jurors about what is the *expected average performance of a team from the midfield of the IYPT competition*
- A significant number of jurors who have spent enough time to get acquainted with the problems and the references about the topics in question, ...

One more element is the *scoresheet* used. It should facilitate coherent grading across all jurors – experienced and newcomers alike and, at the same time, provide information to the teams, what and how is taken into account.

We are all aware that the grading process is notoriously difficult to formalize. Yet, this should not obviate the efforts to introduce improved scoresheets.

The current scoresheet

The official scoresheet used now at IYPT has been introduced in November 2011 (and tweaked a little bit in November 2015). It contains a concise description of the tasks for the *Reporter, Opponent* and *Reviewer*, but most of the space is used for evaluation of the teammates in these three roles.

Jurors are to assign partial grades in three *sections* or, if you prefer, to three *supercategories* – *Physics, Presentation* plus one which is *role-specific (Reporter, Opponent or Reviewer)*. There is a list of items under each *supercategory* which serves to clarify and specify the content of these notions. Partial grades add or deduct points for performance above or below the *expected average performance*.

A lot of jurors are well accustomed to this scoresheet but still, from time to time, one could hear remarks that it is of limited help during the grading process. (Some jurors determine their final grade first, and only afterwards start to wonder what is the best combination of partial grades which would fit it.)

Proposal'16

Last year, a proposal for a new scoresheet was presented and put to trial. Its design was similar to the spirit of the *Factor Analysis* – a relatively small number of *categories* were identified and singled out for evaluation (8 for *Rep*, 4 for *Opp*, 5 for *Rev*). These *categories* had to be more or less exhaustive, so they were somewhat generalized, abstract notions (e.g. '*Correctness*' for *Rep*).

If accepted, this proposal would require specification of a formula (or, rather, a procedure) for calculating the final grade as a function of the partial grades for the different *categories*. This approach was deterministic in the sense that after one grades the categories, the final grade follows automatically. If the jurors were not happy with the outcome they had to modify their partial grades till the desired final grade is reached.

The feedback to this proposal was not conclusive and there were complaints that it would be difficult to reach a clear evaluation of the categories proposed. As the response to this proposal was not overwhelmingly positive it was decided to consider a modification which would put less strain on the jurors.

The concept behind Proposal'17

As a result, a new concept for the scoresheet has been proposed with the explicit intent to make the grading process more 'juror-friendly'. The scoresheet proposed has the following features:

1. Jurors are not asked to evaluate generalized, abstract *categories* (like 'Correctness') but more concrete '*entities*' (like '*time used*' or '*relevant experiments*').
2. Jurors are not asked to evaluate these '*entities*' by means of points, percentages or grades. The assessment is akin to answering *multiple-response questions* as jurors choose between verbal *descriptions* of the degrees of satisfying the expectations for the different entities.
3. There is no automatism in obtaining the partial or final grades – jurors decide how to average or weight their assessments of the '*entities*'.
4. Guidelines are provided about the correspondence between the different levels of performance and the amount of points assigned.
5. Working the scoresheet follows the course of the *Physics Fight* (in theory, there will be no need to go back).

The set of '*entities*' used does not pretend to be exhaustive but it still presents most of the important facets of the teammates' performance. (Adding more '*entities*' runs the risk to include items which are relevant in a small number of cases only and at the same time their presence may urge the teammates to include comments which would not bring extra value to their performance in most of the cases.) Jurors could take into account other considerations, as well, but they should try to keep their scores aligned to the '*performance–grades*' correspondence suggested by the scoresheet.

A more coherent grading and less extreme scores are expected, due to both the detailed description of the degrees of performance and the '*performance–grades*' correspondence suggested.

There is no description of the tasks for the *Reporter*, *Opponent* and *Reviewer* in the scoresheet. This is mainly due to lack of space, but also because it would be more appropriate to have an elaborate explanation elsewhere.