



## **Jury Guidelines 2023**

### **Key Aspects of Uniform Grading**

# 1 OBJECTIVES

In order to ensure a uniform grading for the GYPT tournaments, a physics match is graded according to the standardized jury sheet.

The sheet is displayed on page 10 of this document. It consists of a header (indicating match details such as the teams involved and the name of the juror) and one section each for evaluating the presenting team's as well as the discussing team's performance. At the end of the match both contestants/teams must be awarded a final integer score each from 1 to 10.

The final score is 1 plus the total of the two partial scores for physics (maximum 6 points) and for the fulfillment of the role (maximum 3 points).



match \_\_\_\_\_ problem \_\_\_\_\_ juror \_\_\_\_\_

**PRESENTATION:** \_\_\_\_\_ team \_\_\_\_\_ member \_\_\_\_\_ **Final Grade**

		1 + <input type="text" value="6"/> / 6 <b>Physics</b>				+ <input type="text" value="3"/> / 3 <b>Role</b>			= <input type="text" value=""/>	
+1.0	well understood	very clear	sophisticated	abundant	full model	good match	impressive	assertive	all time used	<input type="text" value=""/> <input type="text" value="G"/> <input type="text" value=""/>
+0.5	considered	explained	sufficient	by results	basic	some	coherent	R/D	fair	
0	disregarded	incomplete	lacking	too few	no predictions	not done	confusing	hesitant	inefficient	
	Task Interpretation	Basic Explanation	Experimental Setup	Experimental Results	Theory/Modelling	Theo ↔ Exp. Comparison	Presentation Style	Discussion Behaviour	Time Management	Personal Impression

**DISCUSSION:** \_\_\_\_\_ team \_\_\_\_\_ member \_\_\_\_\_ **Final Grade**

		1 + <input type="text" value="6"/> / 6 <b>Physics</b>				+ <input type="text" value="3"/> / 3 <b>Role</b>			= <input type="text" value=""/>	
+1.0	deep/detailed	scientific	apt & specific	comprehensive	many correct	improvements	follow-up questions	polite	all time used	<input type="text" value=""/> <input type="text" value="G"/> <input type="text" value=""/>
+0.5	main points	appropriate	interesting	main points	D	limits discussed	reasonable	good	fair	
0	almost nothing	confusing	superficial	fragmentary	none	none	unorganized	poor	inefficient	
	Understanding of Presentation	Argumentation Style	Topics (Quality)	Completeness	Own Opinions	Suggestions	Discussion Structure	Discussion Conduct	Time Management	Personal Impression

**Figure 1.1:** Jury sheet and recommended order of evaluating the partial scores. Criteria indicated by “R” are most likely to be filled during the Report, those indicated by “D” during the discussion. Fields marked by “G” can be evaluated during grading.

## 1 Objectives

time	stage
	introduction (chair)
2 min	preparation of presentation (team A)
12 min	presentation
2 min	preparation of discussion (team B)
12 min	discussion (team A, team B)
4 min	jury questions
	grading
5-10 min	feed-back
<hr/>	
≈ 42 min	

**Table 1.1:** Course of a GYPT Match. The 12 minute presentation by one team is followed by a 12 minute discussion between both teams. After that, the jury is allowed to ask short clarifying questions to both teams before proceeding to grading.

During the course of the match (see match phases in table 1.1), we recommend the partial scores to be evaluated during the relevant phase of the match (see figure 1.1). Criteria indicated by “R” are most likely to be filled during the presentation, those indicated by “D” during the discussion. Fields marked as “G” can be evaluated during grading (after jury questions).

Each criteria is to be evaluated on a decimal scale from 0 to 1 by marking the respective spot on the jury sheet (see page 11 for an example of a filled jury sheet). In order to ease the process of grading, the total of the first six criteria (that make up the score achieved in the physics section) is calculated independent from the total of the latter three criteria (that make up the score achieved in fulfilling the role during presentation or discussion). The final score is then calculated as  $1 + \{\text{physics}\} + \{\text{role}\}$ . Whereas all partial scores are allowed to be a decimal of arbitrary precision, the final score must be rounded to a decimal from 1 to 10. Rounding can be supported by the optional criteria named “Personal Impression” on the right hand side of the sheet.

The jury sheet has meanwhile been used extensively and has some properties one should be aware of when using it.

**Order of Criteria** Presentation/Physics and Discussion/Physics are arranged in a way, that would make the fulfillment of the criteria harder from the left to the right. For the section of presentation, clearly the development of a theoretical model as well as it’s comparison to experimental data is the hardest task of all. We explicitly expect mostly low marks in these sections for most of the presentations. The criteria are important, however, when it comes to distinguishing very good from even better presentations. The same applies for the respective order of criteria in the section of discussion.

**Adaption to Type of Problem/Focus** A structured jury sheet features the major drawback of the lack of being able to evaluate presentations with a strong focus on experiments or theory accordingly. Within the given set of GYPT tasks, however, some kind of differentiation is specifically required. The following two factors should be especially taken into account

1. The complexity of the theoretical model differs significantly between e.g. mechanical and fluid dynamics problems. Please evaluate the theoretical efforts done in the con-

## 1 Objectives

text of what could potentially be done by secondary school students and adapt your demands for full credit accordingly.

2. Regardless of the problem, the presentation may focus on either the experimental or the theoretical aspects of the task. Please take this into consideration by being more generous in e.g. the experimental part if a very complex and perfect theory was presented.

**Trend towards Average Grades** Average total grades of 4-6 seem to be more likely by design of the jury sheet. However, this assumption does not hold true, when looking at the grades of the regional competitions. In order to ensure a fair grading, however, no points may be granted out of pure sympathy and the full range of each criteria is to be used. Especially, there is a strong demand to actually assess a score of 0 points for criteria that are not or almost not present. Why? Because  $9 \cdot 0.2$  points already results a final grade of 3.

**NEW! Problems worked out together** Due to the supervision situation at the locations, several students often work together on the same problem and then split up into different teams. The presentations are then often very similar. In principle, cooperation is desired. BUT cooperation of a significant extent must be made clear in the presentation. The jury evaluation is not to be adjusted purely on the basis of cooperation. If weaknesses in understanding are evident, this is to be assessed anyway. The equality of the presentations should remain within acceptable limits. Exactly the same slides with exactly the same memorised text is not okay. (The supervisors should prevent this in advance.)

**NEW! Citing sources/references in presentations** Participants should include sources and references to content in the presentation on the same slide, not at the end. Obviously missing references are to be evaluated (slightly) negatively in the corresponding sections of the jury sheet.

If the presenter does not provide clear literature references or does not provide sources at all, it is strongly recommended to ask for clarification during the jury questions.

**NEW! Age and work situation of the participants** Jurors often tend to grade younger students or students who have not worked at a GYPT site better. For an uniform evaluation, these preconditions should not be included in the grade. However, they should be emphasised in the oral jury feedback.



Any oral feedback from any juror must also include at least one positive aspect.

The following pages contain specific scoring guidelines for all criteria on the jury sheet.

## 2 SCORING GUIDELINES

### 2.1 Presentation

Generally, the Reporter's appearance is split into two main parts. A 12-minute presentation as well as an up to 12 minute discussion with the opposing team. The Reporter's overall performance is evaluated in 10 different categories of aspects of the presentation as well as the discussion. Each individual criteria adds a maximum of 1 point to the total score.

**Task Interpretation** All content of the presentation should follow the respective GYPT task. The interpretation of the task must be in accordance with the respective operating instruction (*study, investigate, explain,...*) and the presentation should express that the task is **well understood** and interpreted. On the other hand, if the experiment conducted by the reporter is completely different from the one specified in the task, this criteria can be called **disregarded**. If the task is not explicitly named in the presentation, one should not award more than 0.5 points in this section. We call a task **considered** when the task is quoted and the basic execution is in accordance with the task but the execution is oversimplified, or the task requires an explanation but no explanation is given, or anything similar.

**Basic Explanation** Besides the presentation being in accordance with the task, the basic explanation is the most important criteria of evaluating the Reporter's performance. A **very clear** explanation must clearly explain why the respective phenomenon occurs in the exact same way as it is observed, thus answering two questions:

- Why does the phenomenon occur?
- Under what circumstances does the phenomenon occur?
- Why does it happen in exactly this way?

Full credit is assigned, if the basic explanation provides fully qualified answers to all of the questions above without any open questions left to the audience. On the other hand, no or almost no basic/oral explanation on the phenomenon is assigned **incomplete**, 0 points. For anything in between, the respective juror must decide to what amount the observation can be **explained** by the given explanation.

**Experimental Setup** GYPT tasks often strongly focus on the experimental exploration of the physical phenomenon. Therefore, an experimental setup is an essential part of every contribution. The following aspects should be considered/included in a **sophisticated** experimental setup:

- good overall scientific quality of the setup
- reproducibility of measurement results given or proven
- setup should be shown, labeled and explained
- setup should be capable of measuring quantities that are important in the context of the task

## 2 Scoring Guidelines

- all measurement and analysis methods should be clearly explained

Depending on how many of the aspects above are fully or partially answered, a score between 0 and 1 may be assigned to the physical quality of the experimental aspects. If the reporter has not taken any efforts to gain control over one or several experimental parameters (no experimental setup, e.g. wild testing), the report is **lacking** a scientific setup and no credit should be given in this field, respectively.

**Experimental Results** Most of the GYPT tasks require the experimental results to be displayed in graphs under consideration of the known scientific standards (axes labels, units, ...). The experimental results should therefore be judged by both the quantity (number of parameters varied) as well as the quality (mode of presentation). When judging the quantity, jurors should keep in mind the effort required to measure a specific parameter and relate the estimated effort to a student's capabilities. Therefore, a parameter that is very hard to measure may be worth just as much as several **easy** parameters representing the **key results** within the consistent scale of the juror. Apart from numerical parameters, videos may help to understand the experimental findings and should be appreciated in terms of the scoring, if they are relevant. Full credit is commonly assigned, when the reporter shows an **abundant** number of relevant parameters that are systematically measured and displayed at high quality.

**Theory Modeling** The next two criteria are highly dependent on the problem presented by the reporter. Some tasks (e.g. the mechanics problems) allow a rather easy/intuitive theoretical modeling of the experiment and thus, in the best case a **full model** can be expected. Other problems, such as those from the field of fluid dynamics, can be extremely complicated in terms of the **basic** theoretical modeling. In the latter cases, the reporter is free to put special emphasis on the experimental part and reduce the theoretical treatment to **hypotheses and predicted proportionalities**. In any case, a report with **no theoretical predictions/hypotheses** must be awarded 0 points in this criteria. Anything else should be evaluated in accordance with the personal impression of the difficulty of the task (from the theoretical side) under special consideration of the focus of the report. Hence, a report with a strong focus on experiments with a rather hard physical background may still be awarded full credit in this section, if reasonable predictions together with a large number of correct hypothesis or an oversimplified theoretical model are stated. For either case, also the formal aspects of the theoretical solution (explanation of symbols/equations, quotation of literature) may be taken into consideration by the jury.

**Theory ↔ Experiment Comparison** The Theory-Experiment comparison is the most complicated aspect of the evaluation and meant to distinguish high-class presentations from one another. Please be prepared of the comparison to be **not done** by most of the reporters, which means assessing 0 points in this section is completely fine for most of the reports. On the other hand, jurors should value any comparison to any sort of theoretical model/prediction within the limitations stated given by the difficulty of the theoretical model. Top presentations commonly include two theoretical trends/models/predictions with a **good match** to a sufficient amount of experimental data. A theoretical trend is not proven before it is backed up by at least 4 measurement points. Again, special credit may be assigned up to the jurors personal standards if it is especially hard to measure enough data points.

**Presentation Style** The assessment of the presentation style is composed of the common formal demands to a scientific talk. A **coherent** presentation should proceed at a reasonable pace and be held using loud & clear voice. Also an appropriate amount of text on the slides should support the talk of the reporter, rather than the report consisting of a large amount of text,

## 2 Scoring Guidelines

that is either repeated orally or mostly skipped. Additionally, a penalty may be assigned for rushed or **confusing** presentations with either too many slides or content that one can not be followed. If all of these stylistic criteria, however, are fulfilled in an outstanding way, the report may be called **impressive** and awarded full credit.

**Discussion Behavior** The scientific discussion from the reporter's side should express that he/she is **confident** in handling questions related to his/her own work. This means understanding the question posed by the discussion partner on the one hand, as well as stating own, clear and competent scientific answers in a **self-assertive** way on the other hand.

**Time Management** Time management is a major participants' task in the GYPT and should therefore be treated consistently. Any presentation not exceeding the lower limit of **4 minutes** is called **inefficient** and awarded 0 points. A **fair** voting of 0.5 points are given, if the presentation is approximately **9 minutes** long. **All time used** is awarded for presentations of a total length of at least 11:30 minutes. In all cases, the report should not be artificially extended just to meet these criteria. In general only **effectively used** time should award points, roughly 0.1 points to be added for every effective minute between 4 and 9 minutes. If the reporter is forced to stop the presentation by the chair without any conceivable efforts to come to an end before the 12 minute mark, a penalty of 0.25 points is subtracted.

**Personal Impression** This field is a possibility to add/subtract any bonus/minus points, that are not covered by any other aspects of the sheet to/from the reporter's performance. Bonus criteria can be of almost any kinds. Typical examples include an impressive live demonstration of the effect or one/many of the requirements of the jury sheet being fulfilled in an outstanding way, that is way above the standard of the tournament. This field can also be used to include a certain personal aspect when the sum of all partial grades is N.5 before rounding the final grade to an integer of N or N+1. Such aspects could be the juror's impression of the scientific understanding of the reporter (often expressed during jury questions) or if the report put a special focus on theory/experiment in a very competent way and had to slightly neglect some of the other aspects. However, the intention behind this section is **not to make generally good reports even better** or weaker ones even worse. So deciding for plus or minus here cannot be justified by the overall impression of the report, but only by an additional, personal aspect of the respective juror.

### 2.2 Discussion

Throughout all aspects, the discussion lead by the Opponent should generally respond to the solution by the reporter rather than ignoring it in favour of a completely new solution. One significant exception is a missing key aspect in the presentation (e.g. a missing basic explanation), which the opponent may choose to bring in as new material to a reasonable amount. The conduct during the discussion should be fair and constructive. During the discussion, the usage of flip-chart/board and a reasonable amount of help by any of the team members are allowed.

**Understanding of Presentation** One important aspect is the understanding of the presentation by the opponent. A **deep/detailed** understanding of the presentation is often expressed by proper quotations of the statements made during the former rather than just pointing out superficial **main points**. If the Opponent is not able to show any understanding of the reporter's solution/statements, no points should be added in this section.

## 2 Scoring Guidelines

**Argumentation Style** The opponent is asked to pose his/her questions in a **scientific** way so that it can be easily understood by any of the parties involved in the discussion. A **confusing** argumentation style is commonly accompanied by a lot of re-explaining being required in order to make the question understandable to the reporter and/or the jury. Jurors are free to judge the Opponent's argumentation style based on their personal impression along the full scale from 0 to 1 point.

**Topics (Quality)** The scientific topics discussed by the opponent can be of significantly different quality. If only **superficial** and obvious things are discussed, these empty words should not be overestimated. The overall measure of the quality of topics is the physical level and correctness of the statements made by the opponent.

**Completeness** Despite the physical quality of the topics brought up during the discussion, also the overall **completeness** in the context of the report should be evaluated. If all points of discussion are relevant, targeting a large number of different aspects of the report, the discussion is considered **comprehensive** and, hence, full credit can be assigned in this criteria. In case of a weak report, a **comprehensive** discussion should also try to fill in the most important missed aspects of the phenomenon (e.g. a missing basic explanation). Prepared questions, that cause the discussion to be rather **fragmentary** in the context of the solution presented by the reporter, are not the desired style of a discussion and should therefore lead to subtractions in the grading.

**Own Opinions** It is the Opponent's duty to not only bring up certain points for discussion, but also to point out his/her own opinion on the questions asked. Please be prepared that most of the participants will not state any own opinions at all, which means that no points can be added in this section. Likewise, physically wrong statements cannot result in any addition either. Expressing **some** physically correct opinions, however, should lead to an addition to the score in this section up to a full credit of 1 point for **many correct** statements.

**Suggestions** The hardest part of the Opponent's performance is clearly the suggestion of improvements to the Reporter's setup or analysis methods. Any efforts on expressing reasonable **improvements** or pointing out how experimental/theoretical **limitations** could potentially be overcome should be significantly appropriated by the jury. In analogy with the evaluation of the Reporter's performance, the intention behind this section is mainly to distinguish very good from even better Discussions. Most of the discussions are expected to result in a very weak or zero score here.

**Discussion Structure** Perfect discussions follow a clear strategy from the beginning to the end of the 12 minute performance rather than asking questions in a random, **unorganized** way. Another characteristic of high quality discussions is a **reasonable** balancing of mentioning both strong and weak aspects of the report. As a last aspect, jurors should also value if the opponent sticks to a specific topic asking **follow-up questions** rather than just playing a question-answer game.

**Discussion Conduct** The Opponent is asked to lead the discussion in a **polite** way. As scientific discussions among students tend to be rough sometimes, any shouting or other impolite behavior expressed by the Opponent must lead to significant subtractions in this field. In **good** discussions, both parties hear each other out rather than teaching each other a lesson.

**Time Management** Perfect time management (**all time used**) is accredited, if the Opponent manages to finish the discussion within 11:30 to 12 minutes by him-/herself. Of course, full score would also require that the discussion as such was not artificially lengthened by the Opponent



## 2 Scoring Guidelines

for the sole purpose of filling the time limit. In analogy to the Presentation, -0.25 points are subtracted, if the discussion must be stopped by the chair. As with the presentation, 0 points are awarded for discussions of less than 4 minutes and 0.5 points for a discussion of 9 minutes. For any other length, the points should roughly follow +0.1 for every minute between 4 and 9, and +0.1 for every half minute up to 12 min.

**Personal Impression** Again, the juror's personal impression can be used to round the Opponent's performance up to a final integer score. Additionally, the personal impression of the physical understanding may be evaluated in this section as in the respective section of the Reporter accordingly. Third, some credit for extraordinary methods during discussion (such as nice flip-chart drawings or similar), that are not required by any of the other criteria, may be added in this section.

### 2.3 Jury questions

Questions by the jury should mostly target the physical understanding of the participants. Please be reminded, that according to the rules of the tournament any team member is allowed to answer any jury question. The questions (including answers) **must not exceed more than 1 minute** in total and should be clear and easy to understand for the participants. Please also keep in mind, that questions should ask for the actual content of the presentation and discussion rather than completely new or irrelevant side aspects. Common questions are

- Physical explanation & influence of parameters
  - "What would change if...?"
  - key parameter / governing equation
- Limits of the model
  - "Given perfect conditions...maximum efficiency?"

Please avoid to ask questions like

- ~~How long have you been working on the problem?~~
- ~~Why did you only measure three points?~~
- ~~Why didn't you...?~~

The answers to jury questions may be assessed either in the respective criteria on the jury sheet or among "personal impression" if none of the criteria fits.



# GYPT JURY SHEET



## German Young Physicists' Tournament

3 – 5 March 2023 | Bad Honnef

16/2/2023  
13:36

match \_\_\_\_\_ problem \_\_\_\_\_ juror \_\_\_\_\_ team \_\_\_\_\_ member \_\_\_\_\_

### PRESENTATION:

Final Grade

<p><b>1 +</b> <input type="text" value=""/> <b>/6</b> <b>Physics</b></p>		<p><b>+</b> <input type="text" value=""/> <b>/3</b> <b>Role</b></p>		<p><b>=</b> <input type="text" value=""/></p>						
<p><b>+1.0</b> <input type="text" value=""/></p> <p><b>+0.5</b> <input type="text" value=""/></p> <p><b>0</b> <input type="text" value=""/></p>	<p>well understood</p> <p>considered</p> <p>disregarded</p> <p>Task Interpretation</p>	<p>very clear</p> <p>explained</p> <p>incomplete</p> <p>Basic Explanation</p>	<p>sophisticated</p> <p>sufficient</p> <p>lacking</p> <p>Experimental Setup</p>	<p>abundant</p> <p>key results</p> <p>too few</p> <p>Experimental Results</p>	<p>full model</p> <p>basic</p> <p>no predictions</p> <p>Theory/Modelling</p>	<p>good match</p> <p>some</p> <p>not done</p> <p>Theo ↔ Exp. Comparison</p>	<p>impressive</p> <p>coherent</p> <p>confusing</p> <p>Presentation Style</p>	<p>assertive</p> <p>confident</p> <p>hesitant</p> <p>Discussion Behaviour</p>	<p>all time used</p> <p>fair</p> <p>inefficient</p> <p>Time Management</p>	<p><input type="text" value=""/></p> <p><b>+</b></p> <p><b>neutral</b></p> <p><b>-</b></p> <p>Personal Impression</p>

### DISCUSSION:

Final Grade

<p><b>1 +</b> <input type="text" value=""/> <b>/6</b> <b>Physics</b></p>		<p><b>+</b> <input type="text" value=""/> <b>/3</b> <b>Role</b></p>		<p><b>=</b> <input type="text" value=""/></p>						
<p><b>+1.0</b> <input type="text" value=""/></p> <p><b>+0.5</b> <input type="text" value=""/></p> <p><b>0</b> <input type="text" value=""/></p>	<p>deep/detailed</p> <p>main points</p> <p>almost nothing</p> <p>Understanding of Presentation</p>	<p>scientific</p> <p>appropriate</p> <p>confusing</p> <p>Argumentation Style</p>	<p>apt &amp; specific</p> <p>interesting</p> <p>superficial</p> <p>Topics (Quality)</p>	<p>comprehensive</p> <p>main points</p> <p>fragmentary</p> <p>Completeness</p>	<p>many correct</p> <p>some stated</p> <p>none</p> <p>Own Opinions</p>	<p>improvements</p> <p>limits discussed</p> <p>none</p> <p>Suggestions</p>	<p>follow-up questions</p> <p>reasonable</p> <p>unorganized</p> <p>Discussion Structure</p>	<p>polite</p> <p>good</p> <p>poor</p> <p>Discussion Conduct</p>	<p>all time used</p> <p>fair</p> <p>inefficient</p> <p>Time Management</p>	<p><input type="text" value=""/></p> <p><b>+</b></p> <p><b>neutral</b></p> <p><b>-</b></p> <p>Personal Impression</p>



# GYPT JURY SHEET

## German Young Physicists' Tournament

6 – 7 March 2021 | Digital Tournament

12/2/2021  
12:24

match C 13:42 problem 18 Sample Problem juror Perfect Juror

**PRESENTATION: Team A**  First Member  Second Member  Third Member **Final Grade**

1	well understood	very clear	abundant	full model	good match	impressive	impressive	+ 0,5	1,5/3	Role	=	Personal Impression
	considered	explained	key results	basic	some							
	disregarded	incomplete	too few	no predictions	no	confusing	confusing	0				
	Task Interpretation	Basic Explanation	Experimental Setup	Theory/Modelling	Theo ↔ Exp. Comparison	Presentation Style						

**DISCUSSION: Team B**  First Member  Second Member  Third Member **Final Grade**

1	deep/detailed	scientific	comprehensive	many correct	improvements	follow-up questions	reasonable	unorganized	+ 2/3	Role	=	Personal Impression
	main points	appropriate	main points	some stated	limits discussed							
	almost nothing	confusing	fragmentary	none	none	Discussion Structure	Discussion Conduct	poor				
	Understanding of Presentation	Argumentation Style	Completeness	Own Opinions	Suggestions							