Guidance for problem captains

Or: how to write an olympiad rubric

Evan Chen

November 6, 2021

This is a document describing some guidance on how to write a marking scheme for an olympiad-style competition. It was written for the US Ersatz Math Olympiad (https://web.evanchen.cc/usemo.html), but advice here could easily be adopted for other competitions as well.

Examples of past rubrics are provided in the appendix. In addition, the USEMO publishes in full the rubrics for all the problems written every year alongside the solutions and reports. So this can be a good source of examples.

Contents

§1 Philosophy

The point of a rubric is to **give guidelines for the majority of "normal" papers**. This is usually done by having certain milestones or mistakes and the corresponding score to assign for them.

It is *not* the purpose of the rubric to be a deterministic algorithm for assigning a score to *every* single student paper. For most problems, that's simply impossible. When a strange situation comes up — like a completely different solution, or an egregious and outlandish way to screw up — likely your rubric will not cover this case. And that's okay! As long as there are only a few edge cases, the **problem captain** (you) can look at all of them, and make the call for each of them. Since there's one problem captain, that ensures all the off-rubric cases are dealt with uniformly.

So, your goal is to have a reasonable amount of detail, such that for most papers, the graders know what score it obtains. On the other hand, you don't need to bend over backwards reaching for things to add, dreaming up of all sorts of bizarre mistakes or so on. It won't work. You simply need to have enough written to handle "normal" papers.

§2 Numbers

§2.1 Success vs failure

By tradition, olympiad problems are usually graded out of 7 points. Most rubrics differentiate between:

- essentially solved (5-7 points), sometimes called 7^- papers;
- or not solved (0-2 points), sometimes called 0^+ papers.

Scores of 3 and 4 are rare, though they can occur, particularly in weird "off-rubric" cases, or situations where you have no idea how the student didn't finish the problem given what they wrote, but they have indeed not finished. You'd usually use these scores in cases where it's tough to say whether the student has essentially solved the problem or not. Or in two-part problems.

The reason for this tradition is that it implies total score is more or less 7 times number of problem solved, plus or minus some pocket change. That way if a student scores higher than another student by more than a few points, it means that they solved more problems, rather than reflecting some idiosyncrasy of the marking scheme.

§2.2 Additivity

In order to achieve success-vs-failure, rubrics often do NOT use standard addition. Here are a few examples of what I mean.

- Consider a problem that is two parts. In grader jargon, an example rubric for that problem might colloquially be summarized as $2 + 2 = 7$, meaning doing either half of the problem is worth 2 points, getting both halves is worth 7. (On both extremes, I have seen $3 + 5 = 7$ and $1 + 1 = 7$ rubrics as well.)
- On a rubric, there might be three or four different starting observations that are worth 1 or 2 points. Often, a 0^+ solution gets the maximum, rather than the sum, of these little partial bits: $1 \oplus 1 \oplus 1 \oplus 1 = 2$.

  This also applies to deductions. Often, there are a few different kind of mistakes that a 7^- solution could make; one might declare that at most one deduction (the largest) should be applied.

- If you forget to do some easy step (e.g. the base case of an induction), and the rubric penalizes a 7^- solution by 1 point for this, it is not necessarily the case that the rubric would award 1 point for doing that step for a 0^+ solution.

This means that when you set your rubric, it's typically **good practice to specify how items add**. For some rubrics, normal addition makes the most sense; for others, $2 + 2 = 7$ is the right call, etc. You can pick whatever you think makes sense for your problem, just remember to say explicitly.

§2.3 Multiple approaches

Many problems have multiple approaches. Roughly, you want to create a separate rubric for each "class" of approach, where the grouping is problem-dependent and up to you.

Items from different approaches should never be additive (but in some situations it makes sense to have items that appear in all approaches). Graders will usually take the maximum score across all approaches.

§3 Tips for actually sitting down and typing a rubric

Some pointers:

1. Start by reading over the official solutions to the problem, and if available, some examples of student solutions. (If you have a lot of time, better yet, try to solve the problem yourself first before taking any of these steps.) This will get you oriented with the landscape.

2. For each solution, try to map out
   - major milestones of the solution;
   - weaker or different versions of the milestones that you might expect to see, or special cases that students might try;
   - whether any of the claims in the problem might be conjectured without a complete proof (particularly in geometry);
   - common errors or omissions that you expect to see in otherwise complete solutions.

   For each of these, put down a corresponding rubric item.

3. Then, try to choose a point value for each item. This step is delicate, but the easiest way to do it is actually ask your peers! If you are on a good team of graders, there should be a lounge of some sort where people can chat as they get ready, and you can just read the room.

   The reason I suggest this is because it's hard to decide the numbers in a vacuum. Particularly, the correct answer often depends on the *other* rubrics: in a six-problem contest, the generosity of all six rubrics should be roughly in line with each other. So when in doubt, it's often better to discuss numbers. (Arguing with people about how much things should be worth is pretty fun anyhow.)

4. More generally, get as much feedback as you can.

5. For some problems, there will be not many items you can think of, as the problem will feel like it is almost all-or-nothing. That is okay.

§4 Idiosyncrasies

§4.1 Standards across different competitions

The standards for how much partial credit to award incomplete 0^+ solutions, and how strict to be with deductions for small errors, vary dramatically across different competitions and countries. In some sense it is part of the national culture.

For example, the United States is (at least at the time of writing) traditionally one of the most frugal in terms of partial credit.[1] The IMO and Romanian Masters in Math are more generous.

You should talk to the other graders a lot to get a sense of where the individual contest stands!

§4.2 Notes specific to USEMO

A few notes specific to the USEMO.

- Standards are similar to American olympiad (ergo, harsher than IMO). In particular, we're more strict about not letting partial credit build up past 2 points for incomplete solutions than the IMO is, to maintain the distinction between 0^+ and 7^- solutions.
- Incomplete bashes to geometry problems are usually not worth additional partial marks beyond the synthetic observations or milestones.
- Usually special cases aren't worth partial marks unless the ideas used lead to the full solution or the special cases are a substantial "fraction" of all cases.
- We don't typically deduct points for typos and omitted diagrams. (That said, I think if you are a contestant, you should always submit your diagram anyways, no reason not to.)
- We typically try to avoid awarding points for just "name-dropping", and instead require the application. For example, suppose some step of a problem involves applying Z theorem to object X. Then an item like

  - 1 point for mentioning Z theorem

  is *not* desirable; we would much rather have

  - 1 point for applying Z theorem to X or similar, and concluding Y.

  This tries to reward progress towards the problem rather than merely compiling lists of buzz words in solutions.

[1]Here is a long justification for why I (Evan) tend to prefer more strict rubrics, representing my own opinion and not that of the United States. Essentially, I think scoring should follow least surprise and total score should correlate tightly with number of problems solved.

More concretely, I want to avoid a situation where, say, Alice has solved problem 1 for 7 points, and Bob has solved nothing but has some partial marks on problems 1 and 3 that add up to more than 7. Or, Alice has solved problems 1 and 2 completely for 14 points, while Bob has solved problems 1 and 3, but also got a ton of partial marks on problem 2 because the problem captain was a softie, and walks away with 19 points. (I have seen stories like this happen many times.)

In other words, as much as possible, I want the outcome of a competition to not depend much on the marking schemes. This is much easier to achieve if you enforce, across all problems, that incomplete solutions should not be worth too much (e.g. at most 2). This both gives some baseline consistency and ensures addition of partial credits is not likely to approach a full problem.

Of course, the main counterargument against this is to avoid discouraging students with low scores. I can see the sentiment, though I worry about the resulting score inflation.

§A Examples of past USEMO rubrics

These are just for concreteness, and abridged. You can find full rubrics in the past on the USEMO website, we publish the marking scheme together with the results eaech year.

§A.1 Example rubric: USEMO 2019/1 (easy geometry)

> Let $ABCD$ be a cyclic quadrilateral. A circle centered at O passes through B and D and meets lines BA and BC again at points E and F (distinct from A, B, C). Let H denote the orthocenter of triangle DEF. Prove that if lines AC, DO, EF are concurrent, then triangles ABC and EHF are similar.

Most solutions are worth 0 or 7. The following partial items are available but *not additive*:

- 5 points for proving if AC, DO, EF are concurrent implies $AC \perp BD$ but not finishing.
- 4 points for solutions for which spiral similarity justification is entirely absent, but which would be complete if these details were supplied correctly.
- 1 point for proving that $AC \perp BD$ is equivalent to the problem.

There is **no deduction** for configuration issues (such as not using directed angles) or small typos in angle chasing.

No points awarded for noting $\angle ABC = \angle EHF$, or proving/noting that D is the Miquel point of $AEFC$ but not making further progress on the problem.

Computational approaches which are not completed are judged by any geometric content and do not earn other marks.

§A.2 Example rubric: USEMO 2019/2 (medium-hard algebra)

Let $\mathbb{Z}[x]$ denote the set of single-variable polynomials in x with integer coefficients. Find all functions $\theta\colon \mathbb{Z}[x] \to \mathbb{Z}[x]$ (i.e. functions taking polynomials to polynomials) such that

- for any polynomials $p, q \in \mathbb{Z}[x]$, $\theta(p+q) = \theta(p) + \theta(q)$;
- for any polynomial $p \in \mathbb{Z}[x]$, p has an integer root if and only if $\theta(p)$ does.

For solutions which are not complete, the following items are available but *not additive*:

- 0 points for the correct answer.
- 1 point for proving $\theta(1)$ divides $\theta(P)$ over $\mathbb{Q}[x]$
- 1 point for showing $\theta(x)$, $\theta(x^2)$, ... all have a common integer root, or that every pair does.
- 1 point for showing $\theta(x)/\theta(1)$ is a linear polynomial with rational coefficients.
- 2 points are awarded for a solution that starts to make progress on $\theta(x^n)$ for $n \geq 2$, by proving some main lemma or claim. Showing that $\theta(x)$ is a linear multiple of $\theta(1)$ does not earn this point.

  (A common wrong approach is to claim that the rational function $\frac{\theta(x^n)}{\theta(x^{n-1})}$ takes on every integer value; this does not work since $\theta(x^n)$ and $\theta(x^{n-1})$ could have a common root for $n \geq 2$.)

For solutions which are complete with errors, the following deductions apply, and all deductions are additive:

- -1 point for an incorrect answer. This may include forgetting the ± 1, for example.
- -1 point for a minor error. This most commonly applies to students who took the quotient of two integer polynomials and assumed the coefficients were integers when in fact they could be rational numbers, but the solution can be easily patched once this is pointed out.
- -2 points for a more significant error that is easily fixable.

§A.3 Example rubric: USEMO 2020/3 (hard geometry)

Let ABC be an acute triangle with circumcenter O and orthocenter H. Let Γ denote the circumcircle of triangle ABC, and N the midpoint of $\overline{OH}$. The tangents to Γ at B and C, and the line through H perpendicular to line AN, determine a triangle whose circumcircle we denote by ω_A. Define ω_B and ω_C similarly.

Prove that the common chords of ω_A, ω_B, and ω_C are concurrent on line OH.

As usual, incomplete computational approaches earn partial credits only based on the amount of synthetic progress which is made.

No points are awarded for just drawing a diagram or simple observations.

Follow the notation in the typeset official solution. The following rubric items are totally additive:

(a) **1 point** for proving that P and Q are the poles of lines BH_c and CH_b.

(b) **1 point** for proving that T_a is on the radical axis of ω_B and ω_C. This point can be awarded if the proof is conditional on some reasonable description of P and Q, such as (a).

(c) **2 points** for proving that D is on the radical axis of ω_B and ω_C. This point can be awarded if the proof is conditional on some reasonable description of P and Q, such as (a).

(d) **0 points** for commenting that the homothety center of $T_aT_bT_c$ and DEF lies on the line OH.

The four rubric items above, when combined, give a perfect solution worth 7.

If none of the items above are earned: the following rubric item (not additive) is possible:

- 1 point for both claiming that P and Q are the poles of BH_c and CH_b, *and* that the radical axis of ω_B and ω_C is exactly DT_a.

§A.4 Example rubric: USEMO 2021/2 (medium number theory)

Find all integers $n \geq 1$ such that $2^n - 1$ has exactly n positive integer divisors.

In this rubric, none of the items are additive: neither the positive items nor the deductions. Hence an incomplete solution receives the largest positive item, while a complete solution receives 7 minus the largest deduction. Deductions do *not* apply to 0^+ solutions.

Common items for both solutions

- **0 points** for correct solution set.
- **0 points** for proving that special cases of n (odd n, or prime power n) don't work.
- **-1 points** for not mentioning anywhere that all n in the solution set work.
- **-1 points** for a solution which claims that all powers of 2 work (but resolves the other cases correctly). (Stating that it's well-known F_i is prime for $i = 1, \dots, 4$ but not 5, where F_i are the Fermat primes, counts as a correct proof.)
- **-1 points** for a solution which has an incorrect solution set but is otherwise correct (unless the only error is missing $n = 1$, in which case there is no deduction).

First official solution

- **0 points** for just writing down $2^n - 1 = (2^m - 1)(2^m + 1)(2^{2m} + 1) \cdots$.
- **2 points** for proving that one of $2^m - 1$, $2^m + 1$, $2^{2m} + 1$, …is a square.
- **3 points** for proving that $m = 1$ or 3.
- **7 points** for a complete solution.
- **-1 points** for an incorrect proof, or statement without proof, that $2^r - 1$ is only a square when $r = 1$, and/or that $2^r + 1$ is only a square when $r = 3$, if the solution is otherwise correct. (Citing Catalan/Mihailescu counts as a correct proof.)

Second official solution

- **0 points** for just writing down a result of Zsigmondy (that $2^n - 1$ has at least $e_1 + \cdots + e_m$, or $d(n) - 2$, distinct prime factors).
- **2 points** for proving that equality holds in the estimate $s(d(2^n - 1)) \geq \cdots$ or something similar.
- **3 points** for proving n is 6 or a prime power.
- **7 points** for a complete solution.