Self-assessment rubrics to help students develop scientific abilities and instructors to reduce their workload

Sergej Faletič[1], Gorazd Planinšič[1]

[1]*University of Ljubljana Faculty of mathematics and physics, Ljubljana, Slovenia*

Abstract. The course Project lab at the University of Ljubljana, Faculty of mathematics and physics, aims at developing scientific abilities of students in an environment, which simulates an authentic research setting (Planinšič, Eur. J. Phys., 2007). First and second year students in groups of four to five solve open-ended experimental problems and submit a report in the form of a web-page. Students use Scientific Abilities rubrics during their work, and instructors use the same rubrics to evaluate their reports, helping them to improve the reports until they meet agreed criteria. Most of the rubrics are translated Rutgers Scientific Abilities rubrics (Etkina et al. Phys. Rev. ST-PER, 2006), but rubrics for the evaluation of web-reports were developed by us. The use of rubrics significantly reduces the workload on the instructors and at the same time increases the quality of student reports. In the contribution, participants will have an opportunity to familiarize themselves with the rubrics and their use.

Keywords: assessment, physics, project based work, rubrics, self-assessment, teaching.

INTRODUCTION

Laboratory work is central to physics and should be, therefore, central to physics education as well. Otherwise students might get the wrong impression about what physics is [6,7]. Learning only via lectures could give the impression that physics involves finding someone who knows more than you and getting them to pass the knowledge to you. Even traditional laboratory work is usually aimed at ‘proving’ that what the theory says is true. When, in fact, physics is more about discovering new knowledge or applying old knowledge to solve new problems. These two central activities of physicists can be best introduced via project work. But, project work is at least in some part open ended and, therefore, a difficult activity to assess, because the outcome is not entirely predictable. In this context, rather then assessing the outcome, one might try to assess the process that led to the outcome. These are the actual skills that relate to the scientific method and the scientific competences. A way to assess the process is to use rubrics. In the following sections, we describe the course Project laboratory at the University of Ljubljana Faculty of Mathematics and Physics, Slovenia, where we use rubrics. Then we describe the rubrics and their use. In the end, we give a short report on the efficiency of the rubrics compared to other assessment tools that we have used in the past.

PROJECT LABORATORY

"Project laboratory" is a course at the University of Ljubljana Faculty of Mathematics and Physics offered to first and second year students [4]. The course is based on open-ended tasks that help students develop science competences that are also in line with educational recommendations in Europe and the USA [2,3]. Students work on open-ended experimental physics problems. They have three three-hour laboratory sessions over three weeks to solve the problem. After that, the students have another two weeks to prepare and submit a report in the form of a web page. This report is assessed by the teaching assistant and returned to the students to revise and improve. The iterations continue until the report is deemed acceptable.

THE SCIENTIFIC ABILITY RUBRICS

Rubrics are a common tool in the USA, and much research on their use and development can be found in e.g. [8,9,11]. The Scientific Abilities rubrics have been developed at Rutgers University, NJ, USA [1] and are freely available online [5]. The authors of [1] use the term scientific abilities instead of science-process skills to underscore that these are not automatic skills, but are instead processes that students need to use reflectively and critically. Each rubric is a table that represents a broad scientific ability.

RUBRIC B: Ability to design & conduct an observational experiment					
	Scientific Ability	Missing	Inadequate	Needs improvement	Adequate
B1	**Is able to identify the phenomenon to be investigated**	No phenomenon is mentioned.	The description of the phenomenon to be investigated is confusing, or it is not the phenomena of interest.	The description of the phenomenon is vague or incomplete.	The phenomenon to be investigated is clearly stated.
B2	**Is able to design a reliable experiment that investigates the phenomenon**	The experiment does not investigate the phenomenon.	The experiment may not yield any interesting patterns.	Some important aspects of the phenomenon will not be observable.	The experiment might yield interesting patterns relevant to the investigation of the phenomenon.
B3	**Is able to decide what physical quantities are to be measured and identify independent and dependent variables**	The physical quantities are irrelevant.	Only some of physical quantities are relevant.	The physical quantities are relevant. However, independent and dependent variables are not identified.	The physical quantities are relevant and independent and dependent variables are identified.
B4	**Is able to describe how to use available equipment to make measurements**	At least one of the chosen measurements cannot be made with the available equipment.	All chosen measurements can be made, but no details are given about how it is done.	All chosen measurements can be made, but the details of how it is done are vague or incomplete.	All chosen measurements can be made and all details of how it is done are clearly provided.
B5	**Is able to describe what is observed without trying to explain, both in words and by means of a picture of the experimental setup.**	No description is mentioned.	A description is incomplete. No labeled sketch is present. Or, observations are adjusted to fit expectations.	A description is complete, but mixed up with explanations or pattern. The sketch is present but is difficult to understand.	Clearly describes what happens in the experiments both verbally and with a sketch. Provides other representations when necessary (tables and graphs).
B6	**Is able to identify the shortcomings in an experimental and suggest improvements**	No attempt is made to identify any shortcomings of the experimental.	The shortcomings are described vaguely and no suggestions for improvements are made.	Not all aspects of the design are considered in terms of shortcomings or improvements.	All major shortcomings of the experiment are identified and reasonable suggestions for improvement are made.
B7	**Is able to identify a pattern in the data**	No attempt is made to search for a pattern	The pattern described is irrelevant or inconsistent with the data	The pattern has minor errors or omissions. Terms proportional are used without clarity- is the proportionality linear, quadratic, etc.	The patterns represents the relevant trend in the data. When possible, the trend is described in words.
B8	**Is able to represent a pattern mathematically (if applicable)**	No attempt is made to represent a pattern mathematically	The mathematical expression does not represent the trend.	No analysis of how well the expression agrees with the data is included, or some features of the pattern are missing.	The expression represents the trend completely and an analysis of how well it agrees with the data is included.
B9	**Is able to devise an explanation for an observed pattern**	No attempt is made to explain the observed pattern.	An explanation is vague, not testable, or contradicts the pattern.	An explanation contradicts previous knowledge or the reasoning is flawed.	A reasonable explanation is made. It is testable and it explains the observed pattern.

FIGURE 1. A sample of a rubric. This rubric is for an observation experiment. Taken from [5].

The rubric in Figure 1 assesses the ability to design and carry out an observational experiment. The rows represent sub-abilities. In the first column of each row there is the description of the sub-ability. The other columns describe the level to which a sub-ability has been developed. In the cells of the table, there are short descriptions of the activities that represent evidence that the sub-ability has been developed at a particular level. The most important is the last column, which represents the evidence that the ability has been adequately developed. The descriptions in the cells of this column are used as guidelines to describe what is expected of the students.

The descriptions in the other cells are examples of evidence that the sub-ability has not been adequately developed yet. It is impossible to list all the possibilities. In our experience, it is best to list the ones that are encountered most often. These cells should give the students an idea about what could be inadequate, but the students will sometimes need to spot the exact problem themselves.

The Rutgers rubrics use a four level scale to generically describe the level of development of a sub-ability. The first level represents the sub-ability not being present. The students do not know that they are supposed to do what is described in the first column. This is scored 0. The second column represents the level 'inadequate'. Students know that they have to do something related to the sub-ability, but not exactly what. This is scored 1. The third level represents the level 'needs improvement'. Students know what to do, but not exactly how to do it properly. This is scored 2. The last level is 'adequate'. Students know what to do and how to do it. This is scored 3. In the cells in the columns of 1 and 2 points, there are examples of the most common shortcomings that can result in a particular score. Again, it is impossible to list all the possible shortcomings, but what is listed should give the students an idea of what is expected for a particular score.

The rubrics’ main purpose is to be a self-assessment tool. Students are, therefore, presented with the rubrics at the beginning of their project, and are encouraged to consult them often during the project. This use of rubrics is very important, since it gives the students guidance about how to proceed through the project, while still leaving the project open ended. This is very different form any guiding questions that might be used for the purpose of guiding the project. Most questions assume a certain path taken by the students and a certain intermediate outcome. The rubrics do not require any such assumptions.

A typical setting would involve students being given a task, where the wording of the task would describe the phenomenon with the minimum necessary details for the students to be able to reproduce it. The wording would usually end with “investigate how the … depends on the relevant parameters.”

The students would then consult the rubrics. The last column of the rubric in Figure 1, as an example, guides the students to describe the phenomenon, design an apparatus that will enable them to make measurements, determine the independent and dependent variables, be able to describe how to perform the experiment, identify the shortcomings of the set-up and suggest improvements. It further guides them to pay attention to any observed patterns, and attempt to describe them mathematically. In the end, it guides them to suggest an explanation for the phenomenon.

Besides the rubric in Figure 1, there is a rubric, which focuses on the ability to properly analyse data, and another rubric, which focuses on the ability to present the findings. There are also two more rubrics for two more types of experiments. In the context of the Rutgers rubrics, experiments are divided into three types. Observation experiments are used primarily to observe patterns, identify new phenomena, and search for explanations (the rubric in Figure 1). Testing experiments are used to test hypotheses. The best testing experiments are such that different hypotheses predict different outcomes of the same experiment. Based on the outcomes one can then differentiate between the hypotheses. Application experiments are experiments where previous knowledge is applied to solve a problem. For the needs of Project laboratory, we have also developed two more rubrics: a rubric for the ability to write a web report and a rubric for the reflection on the acquired knowledge and skills.

OUR EXPERIENCE WITH THE RUBRICS

In this section, we describe our experience with using the rubrics. We observed three separate changes after their introduction: The workload on the evaluators decreased, the quality of the reports increased, and our experience enabled us to construct and efficiently use our own rubrics for different contexts. Here we discuss the first two and give some notes on what we learned about the development and use of rubrics.

The evaluation efficiency

The introduction of the rubrics reduced the average time of evaluating the report at least by a factor of three. First, the number of words used in a feedback was reduced by a factor of six. Before the rubrics, every shortcoming had to be described and explained, also what is adequate had to be explained in the feedback. With the use of rubrics, only the score on the sub-abilities are given. Students must figure out for themselves what they need to improve. They can ask the evaluator, but before the question, they must explain what they have thought of themselves. Since the reports are returned to students until all the sub-abilities are scored adequate, this usually involves at least three iterations. With the introduction of the rubrics, the number of iterations slightly increased by a factor of 1.7. The total decrease of the workload is then by a factor of three at least.

In terms of the actual time spent on a report we only have an estimate. The reports can be quite extensive, so it takes about 20 minutes to read through them, paying attention to all the figures and equations. This time cannot be reduced. After this, it takes about ten minutes to give the grades on the sub-abilities and maybe another five minutes to write the feedback. So the total time spent on a report can be estimated at about 35 minutes. Previously, again based on our rough estimate, the time was closer to 90 minutes with all the detailed feedback that had to be written. So again, a reduction by a factor of approximately three can be seen. It should be noted that the reduction is smaller for a novice evaluator, before they become familiar with the use of rubrics.

The quality of the reports

The quality of the reports increased with the introduction of the rubrics. To asses this, we randomly selected a number of reports which have already been scored as sufficiently good in previous years by the same evaluator, and they have been evaluated again with the rubrics. Only one of the seven selected reports scored adequate on all sub-abilities. All the rest would have to be returned to the students to be improved, based on the scores on the rubrics.

Notes

We have learned a great deal about how to properly write and use rubrics. First, the Rutgers scientific abilities rubrics are a very good template to start from. They are validated and well-researched. We attempted to shorten the rubrics, but their effect was diminished, therefore we decided to use the original rubrics, translated, instead.

Second, we have learned that it is very important to list the 'adequate' column in great detail, since students use it as guidelines about what needs to be done and also as reference for what has possibly not been done properly, if the score is below adequate. This observation has been noted also in [8].

Third, we realized that it is easier to assess the reports, if the descriptions in all the other columns are not too specific. If they are very specific, one finds oneself checking carefully under which category a specific shortcoming is listed, so that it would not give the wrong feedback to the students. In contrast, a judgment of ‘inadequate’ or ‘needs improvement’ can be made much quicker. However, it is important to list the most common shortcomings that result in a particular competence level. This finding is also noted in [8].

Fourth, we learned that it is very advisable to have training in the use of rubrics for both the students and the evaluators. Someone already familiar with their use should score a report with a novice to show the proper way of scoring. The trend of scores by different evaluators matches well anyway. They generally agree about something not being adequate, but they sometimes disagreed on whether it is inadequate, or needs improvement. Experience from Rutgers University, and elsewhere [1,10,11] is clear that reliability - scores independent of evaluator - can only be achieved through the training of the evaluators.

CONCLUSION

We described a course at the University of Ljubljana Faculty of mathematics and physics, which is based on open-ended project problems for students. We described the use of rubrics to evaluate the students’ reports. We showed that the use of rubrics reduced the workload on the evaluators while improving the quality of the reports. After the course, students are required to write a reflection on what they have learned. They are not specifically asked to write anything about the use of rubrics, but some reflected on this topic, too, and following are some quotes, translated from Slovenian:

"… The rubrics were also useful, because they contain all the important elements of experimental work such as assumptions and experimental uncertainties …"

"… When writing the report we had to consider all the experiments again and all the details that could affect the results, … I learned a lot while thinking about all the parameters, such as control variables and experimental uncertainties, and even more about considering them in general. In doing this the rubrics proved to be very useful …"

"… I think the basic idea of the course is very good, also the srubrics were very carefully designed and proved to be very helpful in our work…"

Based on these spontaneous comments, it appears that the rubrics were appreciated also by the students, because they provide very clear description of teacher expectations and also guidelines for how to meet these expectations, features noted also in [11] as features that contribute to the enhancement of learning.

ACKNOWLEDGEMENT

We would like to acknowledge the help of Eugenia Etkina from Rutgers University, NJ, USA, during the adaptation and translation of the rubrics, and Nejc Davidovič, who developed the rubric for reflection.

REFERENCES

1. Etkina E., Van Heuvelen A., White-Brahmia S., Brookes D. T., Gentile M., Murthy S., Rosengrant D. & Warren A., Scientific abilities and their assessment, *Physical Review Special Topics, Physics Education Research*, *2*, 020103-1 – 020103-15, (2006).
2. European Physical Society. (2009). A European specifications for physics bachelor studies. http://c.ymcdn.com/sites/www.eps.org/resource/resmgr/policy/eps_specification_bphys.pdf (21 may 2016).
3. NGSS Lead States, *Next Generation Science Standards: For States, By States*, National Academies Press, Washington, DC, 2013.
4. Planinšič G., Project laboratory for first-year students, *European Journal of Physics*, *28*, S71 – S82, (2007).
5. Rutgers Physics and Astronomy Education Research group. Scientific abilities. https://sites.google.com/site/scientificabilities/ (22 may 2016).
6. May D. B., Etkina E., College physics students' epistemological self-reflection and its relationship to conceptual learning, *American Journal of Physics*, *70*, 1249 – 1258, (2002).
7. Lising L., Elby A., The impact of epistemology on learning: A case study from introductory physics, *American Journal of Physics*, *73*, 372 – 382, (2005).
8. Mertler C. A., Designing Scoring Rubrics for Your Classroom, *Practical Assessment, Research & Evaluation*, *7*, 1 – 8, (2001).
9. Moskal B. M. & Leydens J. A., Scoring Rubric Development: Validity and Reliability, *Practical Assessment, Research & Evaluation*, *7*, 1 – 6, (2000).
10. Lovorn M. G., Rezaei A. R., Assessing the Assessment: Rubrics Training for Pre-service and New In-service Teachers, *Practical Assessment, Research & Evaluation*, *16*, 1 – 18, (2011).
11. Jonsson A., Svingby G., The use of scoring rubrics: Reliability, validity and educational consequences, *Educational Research Review*, *2*, 130 – 144, (2007).