How to assess experimental skills properly at the physics lab course

[redacted]

July 2, 2014

Contents

1 Introduction

The first year bachelor course 'Natuurkundepracticum' has been improved a lot in the past couple of years. For example, the manual of the course has been made much clearer and much more attention is paid to coaching of the student assistants who guide the students through the course. One of the things which can still be improved is the assessment of the students. The assessment has of course been improved and formed during the years this course exists, however during the meetings at which the final grades are given, it is always hard to justify the grades given on the various experimental skills and communication skills.

This is both because different assistants have different interpretations of these skills and because it is hard to say what exactly is the difference between the two categories in which the skills are divided, as these skill sets tend to be correlated. Also, grades are based on a couple of conversations the assistant had with the student and this is usually not much to hang on to. Hence the assessment of the experimental skills is always subjective.

Therefore, it is a good thing to know as an assistant when to take notes of the performance of students during the lab sessions and to have more or less the same understanding of each skill as the other assistants. The goal of this project for that reason is to propose a model which can help assistants assess students, during the course as well as at the end. The model should make it easier to justify and explain the grades and therefore less subjective.

First the course structure is explained briefly in section 2, and the grading in section 3. It is explained what products are graded and how these are graded at the moment. In section 4 I will address the difference between summative and formative assessment, types of feedback, teaching methods and learning levels for both conceptual and procedural skills. The first three subjects on feedback and teaching will serve as background information on how and when to give proper feedback and what methods of teaching are suitable when. The learning levels will be the crucial tool to assess the experimental skills better, meaning less fingerspitzengefühl is involved.

Based on this a model to assess experimental skills is presented in section 5 which I will apply to different cases to show how this can be used.

2 Structure of the course

An overview of the course, which takes 8 weeks for SBI and MNW students, is given below:

Introduction - 2 weeks

- Introduction experiment(s)
- Introduction to Origin: make graphs and fit data
- Data Analysis

Experiments - 5 to 6 weeks

- Experiment 1: Report with possibility to rewrite it
- Experiment 2: Report and presentation

In the first two weeks the students do either one or two introduction experiments and learn about data analysis. The introduction experiments are so-called "cook book" experiments: Most steps and important questions are in the manual. During the experiments the students learn about Origin, how to make graphs, how to fit data to a model and learn to keep track of a lab journal.

At the end of the introduction the students write a measuring report which should contain the results and a good explanation of those results and some discussion on the reliability and validity of the experiment using data analysis. A lot of feedback and (detailed) comments on the measuring report is given which they can use for reports of the next experiments. No grades will be given for these reports, as these are only for practice.

The rest of the course is for the 'real work' (i.e. the work that will actually be graded). These experiments have an open inquiry format. The students get an introduction on

report 1	20%
report 2	30%
presentation	13%
Experimental skills	18.5%
Communication and insight	18.5%

Table 1: Overview of weights of components of final grade.

the subject of the experiment, but need to find a proper research question, make a work plan and set up the experiment by themselves. The lab assistant will of course help them through this by offering hints, answering and asking questions and giving feedback and reminders.

For the first experiment two writing sessions are planned, one for a first version of the report and one to rewrite the report. For the second experiment the students have one week to write the report but do not get to rewrite it.

Lastly, there will be a presentation on the last experiment the student did. The presentation highlights some part of the experiment and/or an interesting application of the results.

3 Grading

The final grade a student gets at the lab course is mostly determined by the quality of the reports (50% of the grade) and whether the student has good experimental and communication skills and understanding of the subject (35%). The last part of the grade is determined by the presentation. See table 1 for an overview.

3.1 Reports

The reports are graded based on the structure of the report, content of the report and language usage and style in the report. This form needs to be discussed in one of the first meetings of the assistants to make sure every assistant has the same definition of the points on the form in the same manner. Also in this meeting some old graded reports are compared, to see whether all assistants agree on the grade and the feedback given. This way large discrepancies between the grades the different assistants give are prevented.

3.2 Skills

The experimental skills that are assessed are divided in two subsections: 'experimental skills' and 'communication and insight'. The communication skills and insight are necessary to develop and improve experimental skills. The division of the skills into these two groups also makes sure that both kinds of skills are assessed separately.

The experimental skills encompass

- commitment and motivation,
- initiative,
- the ability to work in a structured manner (according to some kind of plan) and

- creativity and originality.

The second group of the skills are communication and insight. These contain for example

- asking (good) questions,
- phrasing problems and questions in an exact and complete way and
- being able to distinguish what is important and what is not.

Through discussions during the coaching sessions it became clear that most assistants and staff members found that this division into two grades is better than one grade for all experimental skills together. If there was only one final grade it would probably always lead to an average grade of around 7/10 for every student as there always good and bad points. One could argue however that the two separate grades given would also average to a 7 and that the two skill sets influence each other greatly. Insight, for example, is needed to conduct an experiment properly, and will influence the work plan and the student's creativity in finding solutions to problems. Asking (the right) questions will in turn influence insight, but initiative is needed to ask those questions in the first place. Also, it is through communication that the skills are assessed.

3.3 Grading of skills

The lab assistants are encouraged to regularly make notes on the performance of students. This can be done in the appropriate cells in the Google Spread Sheet which is used, and should be done (bi)weekly to get a good overview of the student's work attitude. This should be discussed during the coaching sessions to make sure this actually happens.

During the grading session a proper grade can than be found for the experimental skills, insight and communication by discussing each student, using these notes. The hard part of course is convert the comments into a grade. Something that worked out very well in the last semester is that one of the staff members (in this case Jaap Buning) converted the comments into a grade and than the assistants would say whether they agreed on that or whether the grade should be higher or lower. I think we should keep that system, it makes the grades a bit more objective as someone else says something along lines of 'this sounds like an 8' and the assistants can determine whether this grade is 'right'.

4 Assessment of skills

The experimental and communication skills form a large part of the final grade, even though the assessment is very subjective and therefore these skills are the hardest to assess. The research question, work plan, reports and the presentation are much easier to assess as it is much easier to point out which points are not clear or are very good. There is a check list which can be used to evaluate a report or a presentation. It is clear what a good report or presentation should look like, what the introduction, the content and the final conclusions must contain. Also for evaluating the reports and the presentation one doesn't have to differ between assisting and assessing the student as this does not require as much assisting as learning to conduct an experiment properly.

For the experimental skills it is harder to distinguish between assisting and assessing, as the lab assistant has to guide the students to conduct the experiment. Since this is a first year course, most students have just graduated from high school and do not have experience with physics experiments. If they have, the experiments were usually cook book experiments, so they had to follow the recipe to successfully conduct the experiment. This course uses an open inquiry format, so there is no recipe. Guidance is therefore needed to learn the experimental skills such that the student can make up the recipe for the experiment. The same skills are graded at the end of the course.

Two other problems are the fact that students work in pairs and may be difficult to assess them individually and being as objective as possible. Both of these can be partly solved by changing pairs for each experiment and having different assistants for each experiment. We need however more than this to be able to assess the student properly, some procedures that help explain the grade, and to make it clear for the students as well why they got that grade. For example, regularly noting down how the students performed in the Google Spread Sheet (containing grades and assessment) can help greatly as you obviously cannot remember the performance of 20 students during 8 weeks. The problem is however, what things should you note down?

In the next section I will go into the two kinds of assessment there are, formative and summative assessment. In section 4.2 I will go into two possible methods two give feedback and in section 4.4 learning levels for both conceptual and procedural understanding are considered. Lastly, in section 4.3, some teaching methods or levels of tutoring are described.

4.1 Formative and summative assessment

There is no clear definition of formative assessment [1], but it is often contrasted by summative assessment, which is used to evaluate students against some sort of standard. Doing an exam is a form of summative assessment. For evaluating skills there is no such standard, as it is simply not possible to measure someone's ability to communicate. If you would like to have a standard for experimental skills you would need a check list with points like 'the student is able to use the oscilloscope correctly within 10 min'. To fill in such a check list is not doable, since you will have to walk around with it during the lab course, filling it in for 10 students, not having enough time to guide the students as well. Also, it is hard to make a complete check list of all points that the skills compel, and does not take into account someone's development regarding the experimental skills. This is were formative assessment comes in.

Formative assessment, or assessment for learning, is used as a diagnostic tool to check whether teaching has been effective and whether the students have learned. It intends to improve the quality of learning and engages students in the problems and discourse of a given area [2]. This form of assessment is used during the whole lab course, for example when evaluating the measuring reports, the data analysis exercises, the first report and of course during the conversations you have with the students. The last is presumably most important when it comes to teaching experimental skills. Formative assessment, giving feedback on what the students do and encouraging them, is an important part of assisting the student and prepare them to conduct scientific research, the goal of the course. It will help students to improve their experimental, communication and writing skills through the course.

Also, at end of the course, when deciding what grade should be given for the skills, again most of the assessment is formative. The assistants, guided by a staff member (a teacher), discuss the student's experimental and communication skills and a grade is based on this discussion. The grade can be seen as the summative assessment. As there is no absolute end goal, the grade is usually given against the overall performance of the group.

In both formative and summative assessment judgement is involved, but in the case of summative assessment it serves the external world instead of the needs of the student. The grade is there for the record, which is of course important as it shows whether the student is competent. The grade may encourage students and stimulate them, but this is usually only the case for high-achieving students, as students can also be discouraged by lesser grades. Assessment contributes to motivation through recognition of achievement but grading generally is not a motivator [2]. As it is not possible to have a complete list of sub skills and want to take into account whether the student has made progression, the evaluation of the skills therefore remain a bit vague. This should not be considered a problem, but it should however be clear for the student where the grade comes from, we must be able to explain it well.

An absolute grade for performance of experimental skills is not an option, but a list of descriptions and statements regarding these may be of great help grading the skills and explaining the grades. These indicators can be a guide for the assistant, and show which kind of situations and questions to look out for when assessing skills. They may be obvious, but if you have no or little experience assessing experimental skills it is very likely you will not think about it until the very end, when you have to think about a grade. For assistants with experience these may also be of help, as you can use it to justify your grades more easily. Moreover, the assistant has a guide that he or she can bear in mind while discussing problems and progress of the experiment, not only for assessment, but also for guidance. Based on the indicators the assistant can point out what the students should change or keep on doing to improve their experimental skills. The last point will also help students understand their final grades on the experimental skills better.

To summarize: both forms of assessment are important during the lab course, where formative assessment is used during the course to improve teaching and learning how to do scientific research and summative assessment is only done at the end of the course, when it is decided whether the student has developed and performed well.

4.2 Feedback types

When performing formative assessment as described in the last section, the lab assistant is usually giving feedback based on an assessment. According to David Boud and Elizabeth Molloy [3] there are effectively two kinds of feedback, Feedback Mark 1 and Mark 2.

Feedback Mark 1 is based on an idea from the industrial revolution, that an engine or a mechanical system can be regulated through monitoring its output and feeding the output back into the system to control it. In case of human learners, information is given about their current work to influence the quality of subsequent work. To make sure that this information has indeed had some effect in the desired direction, the teacher needs to check this. This constitutes one cycle of feedback. If there has been no effect, no feedback has occurred, so one also needs to make sure that the information or feedback given is actually used.

Of course, one cycle is not always enough; for difficult to attain outcomes (such as understanding of a complex subject) or for less-responsive students more cycles may be needed. The number of cycles depends on a number of variables and is empirical. It depends on whether the goals of the course are realistically set and whether the teacher's advice or interventions are appropriate. So there is also a feedback loop for the learning of teachers so that they improve the quality of their remarks on students' work, which can reduce the number of cycles needed and improve the learning of the students.

The problem with this kind of feedback is that it is not applicable to everyday teaching and learning in higher education. In this case one keeps feeding back information to the learner, until the learner has reached the set goal. As students will graduate at some point, the support for learning will disappear, hence it is necessary to phase out the amount of feedback and guidance. Feedback mark 2 is a model which does take into account these considerations and is therefore also called sustainable feedback.

In the Feedback Mark 1 model learners need someone else to identify and provide information they need to learn, while in a sustainable feedback model, the active role of the learners must also be acknowledged, they are constructors of their own understanding. The feedback model is not mechanistic but responsive, and it becomes a process used by learners to facilitate their learning. This is the kind of feedback also used in the lab course, where feedback is not only given on reports (much like feedback Mark 1), but also during discussions. Both the student as the assistant comment and discuss the progress of the experiment, during which the experimental and communication skills of the students are assessed by giving feedback whereupon the student becomes aware of the quality of their performance. Another important difference with Feedback Mark 1 is that the student is encouraged, and expected, to seek feedback actively instead of waiting for it. Since not all students are prepared to do this, we should help them develop these skills, effectively seeking and utilizing feedback.

During the lab course the students ask for help when they run into experimental problems, difficult to understand theories or writing reports, the assistants help by giving hints, suggestions and answering questions, but also asking questions and giving feedback on their performance. This happens almost automatically as the students do not have a manual in which everything is covered. They have to find information themselves. The assistants do not tell the students how they should do something exactly (only in case of an introductory experiment), they do not have the role of experimental physics 'expert', as in the case of Feedback Mark 1. For example, if the students comes up with a solution of a problem, even if it is not the best solution, the assistant should not tell them that it is not the right solution and tell the students how it should be done. Instead, the assistant should only give feedback and maybe suggest how the solution can be made better if necessary.

4.3 Teaching methods

A way to realise the learning potential of apprenticeship in the cognitive domain was introduced by Collins, Brown and Newman in [4]. This cognitive apprenticeship is a theory of the process where a master of a skill teaches that skill to an apprentice. In this case it encourages students to develop new understandings and to tie this knowledge to meaningful situations. The teacher becomes a guide for learning. Students learn new knowledge and skills by observation and guided supported practice.

There are six teaching methods developed by Collins *et al.* to help students attain cognitive and meta cognitive strategies for using, managing and discovering knowledge. The first three of these, modelling, coaching and scaffolding, are at the core of cognitive apprenticeship and help develop these strategies. Then there two teaching methods, articulation and reflection, which are designed to help students to gain awareness of their problem-solving strategies in comparison to that of experts. The sixth method, exploration, intends to guide the student towards independence. It aims to encourage students to frame and identify problems within the domain on their own. An overview of all six methods is given below.

- **Modelling** A teacher models a new task by explaining exactly what he or she is doing and thinking while performing this task. The student can observe and develop a conceptual model of the processes that are required to accomplish the task.
- **Coaching** The teacher observes students while they carry out a task. The teacher offers hints, scaffolds, gives feedback and reminders, models and provides new tasks aimed at bringing the student closer to expert performance.
- **Scaffolding** In this case the teacher provides support in the form of suggestions, procedural facilitation or physical support. The teacher may have to take over some parts of the overall task which the student is not yet able to do. Scaffolding also requires that the student takes responsibility for the task and make important decisions on their own as soon as possible. This is a phasing out of support, just like feedback is phased out in Feedback Mark 2 described in section 4.2.
- **Articulation** This includes any method of getting students to articulate their knowledge, reasoning or problem-solving processes. This can for example be done by inquiry (the teacher asks questions to allows student to refine and restate previously learned knowledge) or thinking aloud.
- **Reflection** Students are enabled to compare their own problem-solving processes with those of an expert or another student. The goal for a student is to look back on and analyse his or her performance to see how to improve towards expert behaviour.
- **Exploration** Students are given room to solve problems on their own and teach students exploration strategies.

The last three teaching methods are very different of the first three in the sense that the last three methods are tools that can be used while teaching, and the first three are different 'levels' of teaching. Each of these three portrays an 'amount of teaching' needed to teach the student the experimental skills required to conduct an experiment. In the case of modelling everything is explained and demonstrated explicitly and may be necessary when a student is not familiar with the subject of the experiment. In the case of coaching and scaffolding the student gets much more responsibilities and the support provided by the teacher even fades out when the teacher is scaffolding.

The three levels are not necessarily hierarchical. If some mathematical knowledge is required to understand the theory of an experiment but not (yet) known to the students, the teacher has to demonstrate how the theory is applied. However, attaining a graph

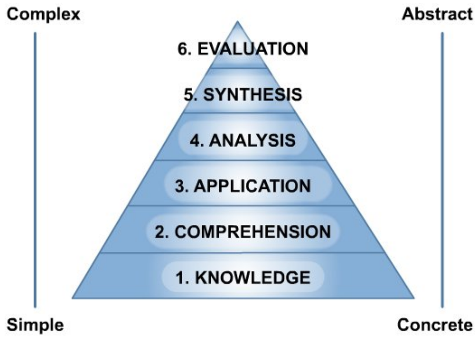

Figure 1: Bloom's pyramid, a classification of learning levels or objectives within education. Simple knowledge-based recall questions and tasks form the base of the pyramid, while increasingly challenging ones form the next levels of comprehension of the subject. Figure taken from [5].

of the results which shows a pattern in the data in a clear way is something that is expected of the student when the teacher is only coaching or scaffolding, depending on the amount of data handling. It really depends on the complexity of the learning goal whether the need to model a task shows whether the student should get a positive or negative assessment on that particular task.

4.4 Learning levels

To assess skills and understanding correctly it is useful to have some classification of learning levels to refer to. A widely used classification of learning levels is Bloom's taxonomy. introduced in 1956, which relates educational objectives to the thinking processes involved.

There are six different levels of cognitive ability in Bloom's taxonomy which are hierarchical as the higher levels subsume the lower levels. The six levels are (lowest to highest, see figure 1): knowledge, comprehension, application, analysis, synthesis and evaluation. In more recent (1986 onwards) literature [6], the highest three levels, analysis, synthesis and evaluation are combined into a fourth level. In the case of science (education) these three tend to be intertwined. In some cases, synthesis is put at the highest level, instead of evaluation. A description for each of the four learning levels for scientific education [5, 6]:

1. The **knowledge and recall** of scientific facts, hypotheses, theories and concepts as well as terminology and convention. The remembering of previously learned material.

2. **Understanding** of scientific knowledge and relationships. This may be shown by expressing information in another form (for example words to numbers) and by explaining, summarizing and interpreting material.

3. **Application** of scientific knowledge and understanding to unfamiliar situations. This implies that the student is able to select from all things learned (such as rules, concepts, principles, theories, etc.) the relevant items and relationships for the new situation. The difference with understanding is that in this case the concepts are applied in a new context.

4. **Analysis** involves the breaking down scientific information into its component parts to understand the structure of the information. This can be done by identification of parts and analysis of the relationships between parts. Reorganising these parts so that a new structure emerges is the **synthesis** of scientific information, and may involve a plan of operation (like a research proposal) or the production of a unique communication (a theme or a speech). The information may also have to be **evaluated** to see whether it is valid, what assumptions are made and what the consequences are.

This taxonomy can be used as description of different levels of conceptual understanding, but it is not necessarily hierarchical, because concepts can become facts with increasing experience. For example, it is not necessary to learn the whole concept of integrating functions and what an anti derivative is every time you want to integrate a function. Therefore application is not always more complex than understanding.

Even though conceptual understanding is required to carry out an experiment successfully, a student also needs experimental skills or procedural understanding to conduct an experiment. For this reason we would also like to have a taxonomy for procedural understanding as well. Richard Gott and Sandra Duggan suggest a procedural taxonomy based on skills and 'concepts of evidence' in [6], just like conceptual understanding is based on knowledge and scientific information or concepts. Concepts of evidence refer to concepts associated with procedural understanding, such as variable identification, accuracy and understanding tables. These concepts are structured around the four main stages of investigative work:

Design of the experiment identification of variables as dependent or independent, fair test (what variables need to be controlled), sample size (how many times should the measurement be done), variable types

Measurement relative scale (choosing sensible values for the dependent variables), range and interval (choosing a sensible range for these values to be able to compare the results to the model), choice of instrument, repeatability of the measurement and accuracy of the measurement

Data handling understanding tables and graph types, patterns (how to represent the behaviour of variables) and multivariate data (how do different variables depend on one another)

Evaluation of the experiment reliability and validity: can the data be believed and can the data answer the research question

The first three stages are often revisited during the experiment. One may find during the measurements that he or she needs a much bigger sample size or note that there is another variable which needs to be controlled to see a pattern. Also, the concepts associated with measurements are about the decisions that have to be made about the measurement and not about the skill of measurement (the ability to use an oscilloscope for example). The last stage, evaluation of the complete experiment, can only be done if one understands the other three stages, since the reliability and validity can only be considered in the context of the strategy of the whole experiment.

The taxonomy for procedural understanding that is suggested in [6] follows in analogy to the conceptual taxonomy:

1. Knowledge and recall of skills
2. Understanding of concepts of evidence
3. Application of concepts of evidence
4. Synthesis of skills and concepts of evidence

This may help in assessing the experimental skills greatly, as this gives a tool to see how complex different skills or concepts are, while the conceptual taxonomy can be used to assess insight. Note that skills in this case do not refer to the experimental skills of section 3.2, they are more basic activities that are necessary, but not sufficient in themselves, to carry out practical work [6]. These skills encompass for example the ability to construct a graph or table or to use a certain measuring device (microscope, oscilloscope, thermometer, ruler). The concepts of evidence are about how and why certain steps should be taken or skills should be used to execute the test.

5 Model for assessment of skills

There two types of skill sets that can be assessed, the conceptual ones or insight and the procedural ones or the experimental skills. Insight has been assessed together with communication skills (see section 3.2), however we also decide whether a student is doing well through communication. This is why the two grades are very correlated. By only assessing insight and experimental skills (and not communication as well) we can remove part of the correlation between the two grades given. Communication is a very important skill and should be improved during the course, but it is not necessary to assess. Asking questions can show initiative and insight. Asking clear questions shows insight, whereas vague questions usually show that a student does not understand the subject good enough.

The first group of skills will thus only be about conceptual understanding, which we can assess using the indicators

- Are the questions asked /problems described clear?
- Do you (assistant) need to ask for more (specific) information to find out what the problem exactly is?
- The student comes up with (good or creative) ideas and input by himself
- Can the student answer questions (clearly) posed during conversations and the presentation?
- The student can relate the theory to the real world experiment[1].
- Negative or null results are understood by the student, for example by finding possible problems regarding the validity of the experiment or caveats in the theory.

[1] Also shows how well the student can evaluate the experiment, whether the results are valid.

The other group of skills that is assessed is the group of experimental skills, which also encompasses initiative, motivation and the ability to work in a structured way and the procedural skills. The last one is actually is part of the stage 'design of the experiment' described in section 4.4. It encompasses the experimental skills which belong to this stage such as understanding and applying the concept of a fair test or deciding what is a good sample size for this experiment. For initiative and motivation indicators are:

- Does the student let you know he/she has a problem when needed (or does the student just wait until you drop by)? (initiative)
- Do the questions the student asks show whether the student does not understand the problem or does not want to understand the problem? (motivation)
- Did the student think about the problem[2] or does he/she just want an answer? (commitment or initiative)
- How does the student cope with negative results? (commitment and initiative)

For assessing the experimental skills one can think of all concepts of evidence mentioned in section 4.4, all skills that are important for the design of the experiment (making the work plan), the measurements, the data handling and the evaluation of the experiment. Each of these have their own indicators, for example, in case of the design of the experiment indicators may be:

- The student is able to see the whole picture, i.e. identify all relevant variables, and does not just focus on one detail or variable. (identification of variables and fair test)
- The student is able to alter a variable in such a way that the other variables stay constant. (fair test)
- Initial results are used to make a (good) work plan. (sample size and fair test)

The list of experimental skills will make it easier to say something about how good or bad the experimental skills are, just by naming which skills were (not) good and why. Assessing whether the skill is 'good' can be done using the taxonomy.

Note that for procedural skills we do expect student not only to understand the concepts, but also that they can apply them. The student needs to be able to actually design a fair test and not only understand the concept of a fair test. For conceptual understanding or insight it depends heavily on the information provided at which level the student should be able to understand it.

The lists of indicators in this section are of course not complete, but can serve a guide and an example of what cases to think of when assessing certain types of skills. One can then assess the skill by using the taxonomies provided in the last section, decide what the learning level of the student is for a couple of skills. Only for 'a couple' of skills, as you will not be able to test the students for each skill. Sometimes there is little to no data analysis necessary, or the repeatability of the experiment had not been discussed, because that was not necessary.

5.1 Cases

This section contains different cases in which students show or don't show certain skills and assessment of these skills and some comments on the case described. In the comments

[2] Can also show insight.

I will try to use the model described in the last section. I will also address what you can do as an assistant when you encounter the situations described. I do not give any grades, as one case is not enough to

Examples 4 and 5 were provided by Leo Polak, 6 - 8 by Matthijs Jansen, 9 by Aernout van der Poel, 10 by Ruud van der Beek.

Example 1 - Force plate

Student A and student B are working on a setup for a spring-mass system, to determine the mass dependence of the damping coefficient. This setup is placed on a force plate to measure the force the spring-mass system applies on it. It is clear to both students that there are some limitations, for example, the force plate cannot measure accurately enough to get a good measurement for masses smaller than ~2 kg. However, when larger masses are put on the spring, the whole setup starts to oscillate horizontally as well. It is very unstable. Also, there are no springs strong enough to use masses of more than ~ 5 kg (as the masses will than lie on the ground).

Their solutions is to shorten the pole they use to hang the spring-mass system on, which indeed makes the system less vulnerable to horizontal movements (these are just smaller), however, the mass of the system now just sits on the force plate, since it is to heavy for the spring.

The students in this case were very motivated to do this experiment, and showed some insight as well: the problem was the instability of the setup which they solved, however they knew that the mass would than just sit on the ground as they knew the springs available were not strong enough for the large masses needed. Therefore their insight was OK, however the adjustment to the setup was not thought through properly which can be reflected in experimental skills (structured working, may also be seen as a lack of insight as well).

Comments

Here, the more general problem is that the students focused too much on one of the variables of their setup, forgetting the other variables (the mass and spring constant). Even though they figured out themselves or have been told that the heavy mass and low spring constant combination may be a problem, they forget about it and focus only on the horizontal oscillations of the setup. There is a good understanding of relative scale at which to measure (the masses should be heavy), but the application fails, as not all variables where identified.

This is seen more often during the lab course, many students make the mistake to focus too much on one variable. Therefore seems to be a skill the students simply need to learn during the course. Also, it is perfectly fine to just try these things out during the development the work plan, it should than be included and worked out in the work plan. This way, the student can get a better feeling for all the different variables and design a fair test, and find out how can the independent variables can be controlled.

The assistant can point out that you can figure these things out before even trying to adjust your setup, and how to see these things, i.e. the assistant needs to model instead of coach. A student can than use this feedback for the next experiment. It is important to make note of these kind of things in the Google sheet so the lab assistant of the next practical can take this into account when assessing the student's (development of) experimental skills.

Example 2 - general
Two students are working on their work plan and when the assistant wants to discuss it, they do not yet have a complete work plan, because the research question is missing. Their first question is 'What actually is the research question?'

Comments
In this case it is clear the question posed is not a 'good' question. It shows a lack of initiative, as they did not already think for themselves about a research question (like 'how to improve the fuel cell', not a good research question, but it is something). So two indicators in this case are:

- Do the questions the student asks show whether the student does not understand the problem or does not want to understand the problem?
- Did the student think about the problem or does he/she just want an answer?

The first could of course also be asked when assessing a student's insight. In this example, however, the question asked by the student does not tell anything about the student's insight, as it is a very general question. Also, what is a 'good' question? Two indicators can be

- Is the question asked/problem described clear?
- Do you (the assistant) need to ask for more (specific) information to find out what the problem exactly is?

These show insight, when someone has good understanding of the subject he can ask much clearer and 'smarter' questions.

Where does such a question come from? The student may just be lazy or has no motivation to do this experiment, but he might as well not have understood the introduction very well. In both cases you need to answer this 'lazy' questions with another question. In this case probably a good answer is 'what is the subject of your experiment?' or 'what do you want to research?' and let them try to capture this in a research question. Afterwards the assistant can come back to discuss whatever research question they came up with. During this conversation one can then find out whether there is a lack of initiative (think about it yourself first), motivation or insight.

Example 3 - general
A student is working on his experiment and seems to be okay. The assistant is busy helping other students. When the assistant asks the student how everything is going, it turns out that he actually has been stuck on something for quite some time.

Comments
Here, the student should have gotten the attention of the assistant earlier; it is not a good idea to just wait for the assistant to drop by and then ask your question or to keep on trying to figure something out when you're clearly missing some information. Especially if you only have 3 days to work on the experiment.

In the first case someone is just lacking some initiative. In the second case it is just stubbornness (just keep on doing something which clearly is not working). In both cases the student just should have asked a question at some point, indicating either lack of initiative or communication skills. The assistant should make clear that they should really just ask if something is not clear and make note of it so you keep in mind you probably should walk by a couple of more times to see whether they're stuck again. Not only is it important to have communication skills to ask for the help you need, but without it the assistants cannot assess the experimental skills properly.

Example 4 - Fuel cell

During the first day two students had been sitting doing almost nothing for quite some time until I came by to get them started. I told them that they should have started trying to get the thing running by themselves, they said they didn't know how.

On the second day they were trying to determine the total efficiency of a hydrogen fuel cell. This requires a measurement of the amount of H_2 that has gone out of the storage cylinder during a certain amount of time while measuring the power supplied by the fuel cell. During the measurement the power can start to plummet, which makes the measurement less useful.

In their setup the power was each time decreasing to less than half the starting value, but the students just decided to keep track of the plummeting power. And continue the way they were going. When I asked them how they were going to determine the efficiency they said they were going to take the average power in the efficiency calculation, but when I asked how, since the decrease was not linear, they didn't know, and when I asked what error they would take for this average power they had no idea.

They were also applying different resistances so they could assess the efficiency for different output powers, but since the power was constantly changing this idea had become kind of pointless. They did not come up with any of the possible ideas to make something out of the experiment like using smaller measurement times, check for which resistances the system remains stable the longest, do some serious flushing before each measurement to get rid of some of the moisture in the cell, etc.

I was a bit fed up with their lack of input and ideas, so I just gave them the ultimate solution, which is to hook up a second electrolyser to the O_2 side of the fuel cell, keeping the O_2 outlet open and create a constant flow, while keeping the H_2 side closed for volume measurement. This helps tremendously since the problem is almost entirely caused by the O_2 side. I really hope they give a proper description of this in their reports, because their initiative during the experiments was not very good, except for asking me a lot of questions.

Comments

We see that on the first day the same problem occurs as in example 3, and it shows a lack of initiative. They should have tried to ask the assistant, and before that try to look up how to start in the literature mentioned in the introduction (if mentioned, but I'll assume it was). Although it is not obvious for all students who are just graduated from high school to look these things up themselves; they are not used to looking the information up themselves as it is usually just given.

Telling that they do not understand what to do is sort of OK during the first day of the first practical. They do need to learn to look for info themselves, so the assistant usually can answer with a simple 'have you looked at ... ?' or 'what do you think what needs to be done to get ... ?' . The fact that they just sat there doing nothing, not only shows the lack of initiative, but also motivation.

The second part can mean two things, the one pointed out already (motivational problems), but also that they just did not understand the subject very well. The assistant should at some point try to explain some basic ideas and give some hints which may help the student think of what the problem is. If they do not understand the subject try to let them first read the literature and then try to figure it out together in a discussion instead of giving the solution. This takes a lot of time of course, which is why giving the solution at some point is just necessary.

It should be noted however that only asking a lot of questions and not trying to think of something yourself does not necessarily mean that your motivation/initiative is the whole problem. As pointed out before, the students still need to learn to look for and interpret information themselves. The lack of input of ideas can also shows that their insight was not good as well, but this can only be found out from discussions.

Some experimental skills that can be assessed here are about measuring: choosing sensible values for the resistances, times etc., what interval should be chosen to keep the system stable enough to get valid and reliable results.

Example 5 - Fuel cell

A student who was trying to determine the total efficiency of a fuel cell was measuring the time it took until some chosen, fixed volume of H_2 had been used. He came to me asking how he should determine the uncertainty in the time measurement. (By now I think it is better, or at least easier for analysis to take a fixed time interval and measure the used volume). He had already come up with some idea by himself which was to take the uncertainty of the volume reading and divide it by the rate of volume decrease. I think this was the correct idea and I really liked the initiative to come up with such a solution by himself. The propagation of the error is then a bit harder because the volume and time error are not independent, but I think that is correct.

Comments

Here, the student shows a multitude of skills. He thinks about value of the result during the measurements, and finds that calculating the uncertainty in time measurement is not trivial. He comes up with an idea which shows insight in the physics and of the mathematics involved. More specifically, he shows how to handle multivariate data (how does time depend on the volume and the rate of volume change). Probably, his communication skills are good as well, as he came up with an idea and asked the assistant what to think of it after explaining it. He shows he can explain his idea and applies it correctly so this shows a high level of understanding the concepts involved.

Also, the assistant has an idea that is presumably better than the idea the student came up with, as it is easier to measure, however, the assistant does not say that the student is wrong. Instead, as an assistant you should only agree with the student that this is indeed correct, assess the student's insight, motivation and initiative positively. Later, maybe during some conversation on the report, you can explain about the other possible solution to the problem. Trying to 'correct' ideas that are OK or explain how it could be done differently right away can also create confusion. This is exactly what feedback Mark 2 (section 4.2) is; the teacher acts as a learning expert instead of 'the' expert.

Example 6 - Fuel cell
The students had decided to find the efficiency of the fuel cell when it is operating at maximum useful power output. To do this, they first measured a PR curve to find at which load the fuel cell gives maximum power output (say 0.4 Ohm). Then they did a few more measurements at this resistance, while also measuring the H_2 usage, effectively measuring the efficiency of the fuel cell.

A few things went wrong here: They didn't take into account the internal resistance of the current meter, so they measured at the wrong resistance. More importantly, there was a large difference in the circumstances in which they measured. In the first part, they had a constant flow of hydrogen, while in the second part, the gas compartments were closed to allow for measurement of the amount of gas used. This strongly influences the working of the fuel cell, so it changed the position of the maximum power.

The total result was that they did not have the results they wanted to have. They concluded that they had failed the experiment, so there could be no conclusion. As the assistant, I tried to convince them that they still could get some reasonable conclusions, and that they also could spend a large part of the discussion on why they measured these values, on what they think caused these strange results, and why the circumstances of the fuel cell have such a large influence.

The students took a real SBI-stance and did not want to go into the experimental details, and were only interested in the maximum power.

I think that this should be reflected in the mark for experimental skills and insight.

Comments
Here, the students did not take into account the change of or existence several variables which should be constant during the experiment, like in the first case, only they did not experience this during the development of the work plan but during the actual measurements. I think it correct to reflect this as bad experimental skills; more specifically choosing sensible values for the resistors, because they did not identify the internal resistance as an important variable of the system. They also did not take into account the environment, so the design of the experiment seems to be a large part of the problem.

It is not clear from this case why this should be a lack of insight, but that conclusion may be something found from discussions with the students. Not being interested in the experiment can show a lack of motivation, even though null results are not as interesting as maximal power, the course is about experimenting.

It happens quite often that an experiment fails, because the experiment was not designed properly or because the theory is not applicable to the research question. It is therefore important to help the students see that this does not mean you can't write a good report and there are still a lot of conclusions to make based on a good discussion. The assistant does have a important role here, guiding the students, let them see during a discussion that a 'failed' experiment does not mean you failed. In this case a good discussion should reflect the student's understanding of a valid and reliable experiment. If they do understand this it can show insight.

During the discussions during the course as well as from the discussion in the reports you can find out whether the student understood the problems, and is able to connect the theory with the real world.

Example 7 - general
Bij de presentaties kwam er achteraf nog een vraag over het kwadratenverband (dat de intensiteit met r^2 afvalt achter een lamp). Daar bleek dat zij (de presentator) dit zelf ook niet goed begreep. In de presentatie gaf ze hier maar kort aandacht aan. Ze kon het niet beter uitleggen dan de formule laten zien. (Na een uitleg van mij begreep men het beter).

Dit lijkt me een duidelijk geval van matig inzicht.

Comments
The presenter either did not put enough thought to the theory of the experiment or simply took the formula for granted without understanding the physics behind it; it does exist and can be used in the following situation. She recalls the knowledge, but does not understand it as she cannot explain it. Not understanding the theory can indeed be a sign of poor insight.

It seems that during the experiment both the student as the assistant did not note that she did not really understand this inverse-square law. Although this seems not plausible as the experiment involved will most likely not get you measurements that follow this law exactly, so the physics behind the formula has probably been discussed while the results were progressed. The law should have been applied to the results, which in practice means to connect the results to the theory.

The presenter can also have problems answering the question because of a black out or nervousness. She may also have underestimated the importance of the subject for the presentation and did not think about it any more. So it could also have to do with presentation skills and these are reflected in the grade for the presentation.

The only thing which you can do when something like this pops up during a presentation is trying to explain this, so the student asking the question gets a satisfactory answer and the presenter can maybe still fix this in the report. It is however not a good example of lacking insight of the formula, as there is much more to it. It can however be put on the table during the grading session, and serve as feedback.

Example 8 - PV
De studenten waren bezig met een PR curve te meten van de zonnecel. Omdat de ene student bij een vorig experiment merkte dat hij te weinig punten bij het maximum had, waren ze nu extra veel metingen aan het doen bij lage weerstand, om dit probleem te voorkomen. Toen later bleek dat het tegenovergestelde gebeurde (te weinig punten bij de andere weerstanden) hebben ze hard doorgemeten om dat op te lossen.

Er was dus sprake van een goede motivatie, maar de experimentele vaardigheid was niet zo goed. Het is handiger om eerst grof te meten, en daarna toe te spitsen.

Comments
The students are motivated to get good results, better than their last results, where they found that they did not have enough data points around the maximal power. So they show that they understand the experimental skill choosing the right interval, which values

should be measured. Application fails a bit as they want to make sure that they do have enough data points at that power, however forget to think about the whole picture.

They probably did not have a good work plan for the first measurements, and tried to improve the plan using the (initial) results showing motivation. They do not seem to understand the pattern of the data, in the sense that you need more than data around the peak of the graph to fit the results to the theory.

Example 9 - Wind mill

Twee studenten voor het tweede experiment windmolen. Ze hebben maar ternauwernood drie meetseries kunnen doen, waar 4 meetseries normaal is, en sommigen er 5 hebben. Het schoot echt niet op met ze. Hoe dat kwam werd mij eigenlijk niet duidelijk. Een halve meetserie hadden ze de frequentie van de windmolen niet goed afgelezen (details weet ik niet meer). Ze zaten echter niet te geiten of iets dergelijks, ze waren gewoon heel langzaam! Gingen niet efficient met de tijd om: meetten frequentie en spanning achter elkaar terwijl dat tegelijk kan, veel dode tijd tussen metingen, naar vorige metingen kijken zonder eerst de volgende te laten lopen etc.

Ik zou dit zien als een gebrek aan initiatief. Hun inzicht was niet bijzonder (slecht noch goed), hoewel ze in discussies ook wat traag waren, moeilijk in 'beweging' te krijgen.

Comments

Planning or making a good work plan seems to be problem here, as they were so slow. Maybe they did not realise that the frequency and voltage could be measured simultaneously, they probably understand how to perform the experiment (as they seem to conduct the experiment correctly from this text), but they do not apply this knowledge well.

If the assistant notices, however, that the students are not going fast enough, because of known reasons (like checking results in-between measurements while it is possible to let one person check the measurement and let the other measure), he probably needs to intervene. Just make a note, but making sure that they have enough data. Of course, there are at least 4 other student groups to coach so this can still happen.

Initiative might be a problem here, which could explain their participation in discussions as well. I would however also think that experimental skills also were not very good, as they could perform the experiment but not in a 'smart' way.

Example 10 - Spectroscopy

De studenten moesten licht bij hele lage intensiteiten meten en hadden niet zelf bedacht strooilicht van andere bronnen tegen te gaan. Toen ik dat voorstelde, wisten ze niet wat ze moesten doen: een lampje uit en een schermpje voor de witte muur zetten die veel licht reflecteerde. Dit moest ik echter helemaal gaan voorzeggen. Gebrek aan creativiteit en begrip.

Comments

The students do not understand that low intensities can only be measured in a (very) dark room and how to achieve this in their case. The problem indeed is that the concept of low intensity and how easily noise can ruin a small signal is not known or understood by them. So insight is low in this case.

When explained what the problem was they did not know how to solve the problem. That could be lack of creativity, but can also just be caused by the fact that the students simply did not understand the problem good enough. This is not clear from the text.

Additionally, they do not think about the fact that there is noise from other light sources, so their results would not be useful if they would go on with the measurement. So (some) experimental skills are also not good.

6 Conclusions

In section 5 a model for assessing experimental skills has been presented and applied to a couple of cases provided by several student assistants. The experimental skills are still divided into two sections, one that contains the conceptual understanding, and one that encompasses the procedural understanding, initiative and motivation. Communication skills are eliminated from the list as we already sue communication as a means to assess and to coach students.

Insight in this case is only about understanding, applying and evaluating scientific information, the procedural understanding is the understanding, application and creation of concepts that are relevant for investigation. These skills are about understanding how and why certain steps need to be taken to get valid and reliable data. Examples of these were pointed out in the case studies.

The subjectivity of the assessment can be taken away partly by explaining grades properly. As there is a list of specific experimental skills, the assistant can hopefully give a better grade. It can be justified by giving examples of the these experimental skills, and explaining whether these skills were good: could the student understand and apply the skill? Additionally, one can also take into account how much coaching the student needed to reach the required level.

The model can also be used to perform formative assessment throughout the course. The assistant not only assesses students, but guides them as well, so feedback is not only given at the end (the justification or explanation of the grade), but also during the course. During discussions it should be clear for the student what they are doing well and what can be improved.

The model needs to be tested during the next courses on physics practicals to find out whether it indeed can help lab assistants assess experimental skills better. During coaching sessions the idea can be explained by using the case studies, but also by letting student assistant bring in cases themselves and discussing them. This can also help to make sure all assistants have the same understanding of the skills. Hopefully this will help assessment of skills during the practicals easier for the assistants and better to understand for the students.

References

[1] Paul Black and Dylan William. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1):7–74, 1998.

[2] David Boud. Assessment and the promotion of academic values. *Studies in Higher Education*, 15(1):101–111, 1990.

[3] David Boud and Elizabeth Molloy. Rethinking models of feedback for learning: the challenge of design. *Studies in Higher Education*, 38(6):698–712, 2013.

[4] Allan Collins, John Seely Brown, and Susan E. Newman. Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In Lauren B. Resnick, editor, *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, pages 453–494. Lawrence Erlbaum Associates, 1989.

[5] Monash University DELTA. Learning levels. `http://courseware.monash.edu.au/DELTA/blooms_pyramid.html`, April 2014.

[6] Richard Gott and Sandra Duggan. *Investigative Work in the Science Curriculum*, chapter Alternative perspectives. Open University press, Buckingham, 1995.