Promoting formative assessment in high school teaching of Physics

Clemens Wagner and Andreas Vaterlaus
Solid State Physics and Education, ETH Zürich, Schafmattstr 16, 8093 Zürich, Switzerland.

E-mail: clemens.wagner@phys.ethz.ch

Abstract

There is ample evidence that formative assessment is one of the most promising teaching tools to improve learning. Due to its broad definition it has to be adapted to high school teaching. We have identified four lines of formative assessment, which should be applied in physics teaching. First, we use formative assessment to monitor student learning, second, students use formative assessment to determine the difference between the learning goals and their actual performance, third, we use diagnostic tools to verify students' understanding and fourth, summative exams are used in a formative way. In the second part of the paper we use the force concept inventory (FCI) as diagnostic tool to show how it can be used as feedback for the students and for the teacher. Moreover, we identified four classes of students in the results of a post-test FCI who appear to be the outcome of the predominant teaching style in Switzerland fostering numerical problem solving rather than conceptual understanding.

Keywords: Formative assessment, diagnostic tools, conceptual change.

Resumen

Se tiene fuerte evidencia que la evaluación formativa es una de las herramientas de enseñanza más prometedoras para mejorar el aprendizaje. Debido a su amplia definición tiene que ser adaptada a la enseñanza secundaria. Aquí hemos identificado cuatro líneas de la evaluación formativa, que deberían aplicarse en la enseñanza de la física. En primer lugar, se puede utilizar la evaluación formativa para monitorear el aprendizaje del estudiante, en segundo lugar, los estudiantes pueden usarla para determinar la diferencia entre las metas del aprendizaje y su rendimiento real. En tercer lugar, se utilizan las herramientas de diagnóstico para verificar la comprensión de los estudiantes y finalmente se pueden utilizar los exámenes sumativos de manera formativa. En la segunda parte del trabajo se usa el "force concept inventory" como herramienta de diagnóstico para mostrar su utilización como retroalimentación para los estudiantes y el profesor. Por otra parte, se identificaron cuatro clases de estudiantes en los resultados de un post-test, que parecen ser un resultado del estilo de enseñanza predominante en Suiza, el cual fomenta la resolución de problemas numéricos en lugar de la comprensión conceptual.

Palabras clave: Evaluación formativa, herramientas de diagnóstico, cambio conceptual.

I. INTRODUCTION

Since the seminal paper by Black and Wiliam [1] formative assessment has been recognized as one of the most promising tools to promote learning. Although some approaches have been proposed to apply formative assessment in physics [2], a clear concept at the high school level is still missing.

The starting point to develop a formative assessment framework is the following inherent and fundamental problem of teaching: Whatever we teach and independent of how we teach we don't know what happens in students' minds. Even in the case of an excellent teaching style and with intelligent and willing students, it might happen that what was understood by the students doesn't correspond to what was taught [3]. As a consequence teaching has, first to

induce student's construction of knowledge, and second to make students' thinking visible. Formative assessment can be used for both. On the one hand the construction of knowledge can be induced using learning activities, which foster critical thinking and explore physical concepts. The clicker system suggested by Mazur [4] represents a possible teaching style to get students involved. On the other hand formative assessment can be used to determine the difference between leaning and learning goals. Note, that formative assessment is not a tool for the teacher to measure students performance (students get no marks for formative assessment tasks) rather it gives students the opportunity to explore their understanding. However, the degree of understanding has to be monitored. We think it is important that students self assess their level of understanding to get aware of their learning deficits.

Besides the monitoring tool and classroom activities a diagnostic tool has to be developed. It provides a detailed feedback to the students about the concepts they got right but also about the presence of misconceptions and wrong beliefs. Furthermore, it probes students' self-assessment and helps to improve it, if necessary. The feedback for the teacher is a summary of the class performance. It can be used to adjust and improve teaching. Finally, summative tests can be used in a formative way. They should confirm the results obtained by the self-assessment tasks. Differences have to be carefully studied.

We are going to conduct a study where 15 classes are taught in a traditional fashion and compare them to 15 classes where formative assessment is used. The instruction period lasts about 30 lessons and covers a basic mechanics course.

A second line of our research focuses on the acquisition of physical concepts by the students. Therefore we have to measure students' understanding in a time dependent fashion. Since some of the formative assessment tests are using a clicker system, which stores the responses in a database, they provide the first set of data. Furthermore, the results of the diagnostic tool are also available on our server system. Finally, the outcome of the pre- and post-test administrated to the students is also analyzed electronically. Taken together, we obtain a whole set of time dependent data for each student. If we determine the misconceptions at the beginning of the course we can follow the conceptual changes of each single student in a detailed manner. Using latent transition analysis we can search for different classes showing the same temporal behavior of the acquisition of concepts or of conceptual change.

Therefore the present paper first highlights our framework of formative assessment in high school teaching of physics and then shows what kind of classes might appear related to the acquisition of physics concepts.

II. GENERAL FRAMEWORK OF FORMATIVE ASSESSMENT IN HIGH SCHOOL TEACHING

Black and Wiliam [5] state that substantial learning gains can be achieved in a classroom setting where formative assessment with feedback to the learner is used in combination with systematic remedial learning work of the students themselves. From the perspective of the learner the introduction of formative assessment with feedback shifts the emphasis from grading towards understanding. When formative assessment with feedback is used, learning deficits are detected and communicated on a regular basis. "Passive" learners would constantly be reminded of their lack in understanding. As we have already mentioned formative assessment can be used by teachers to improve teaching instructions. If no student fulfills the criteria for success one has either to skip that part or to approach the topic from a different perspective. Based on the results of formative assessments, lessons have to be planned where students can remedy their lack of knowledge. In order to

activate students as owners of their own learning they have to take the decision what gap should be closed during these lessons.

In order to apply formative assessment in high school teaching of physics we present a general framework. It can be divided in four different instruments: Monitoring, formative activities, diagnostic tools, and formative use of summative assessments (see Fig. 1).
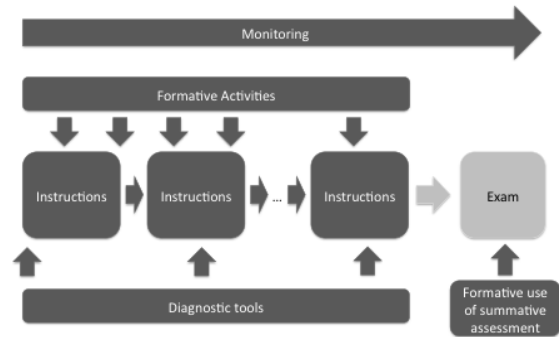

FIGURE 1. Framework of formative assessment. It consists of four tools: Monitoring students' learning, formative activities which induce active processing of information, diagnostic tools to evaluate the level of understanding and the formative use of summative assessment.

Monitoring students' learning requires that students are familiar with the learning objectives. They also need to assess their own state of learning and they must be able to properly determine the difference between the learning goals and their actual knowledge. For the students the monitoring procedure has to be easy to handle for the students. We think of an assessment booklet, which lists the classroom activities and includes learning objectives, rubrics or competence grids in order to determine the learning deficits.

During lessons formative activities are introduced so that student can check their understanding. These are in general short activities and encompass for example

- Quiz.
- "String and sticky tape" or "hands-on" experiments.
- Predict observe explain (POE) experiments (movies or real time).
- Clicker system.

A whole bunch of further activities was suggested by Keeley [6].

However, if we want to test problem-solving strategies or the embedding of knowledge in a larger context one can use activities like

- Jigsaw method.
- Concept map.

All these activities have to be organized in such a way that effective classroom discussions are induced enhancing the

learning effect. It is also clear that the efficiency of learning strongly depends on the quality of questions and problems.

Introducing formative activities during lessons does not only help students to test their understanding but also changes the teaching style. It will be more dynamic whilst alternating between instructions and student activities.

In order to provide substantial and qualified feedback to students we think that it is also important to have a diagnostic tool. It should be designed to probe the understanding of basic principles and concepts. These tests provide individual feedback to the students and "global" feedback to the teacher. Every student gets a detailed response about the concepts he has understood and about the misconceptions he still carries along.

Each diagnostic test is followed by a reflective lesson where the students are given time to learn the proper concept. This requires a flexible planning of the teaching unit [7]. Depending on the outcome of the diagnostic test, the class might be divided into different groups according to the differences in learning deficits. Thus, for each group the learning instructions or material has to be prepared separately. In addition, we suggest to reflect and to discuss the learning strategies. How the assessments are distributed among the lesson plan is presented in Fig. 2.
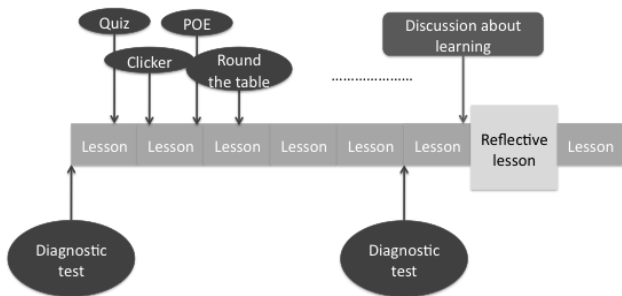

FIGURE 2. Time dependence of using formative assessment tools in a sequence of lessons.

Formative use of summative assessment then focuses on problem solving strategies and how to deal with numerical problems. It should also be a time point to reconsider the self-assessments of the students given for their activities during lessons. If the self-assessment does not predict the outcome of the summative exam the origin of the difference has to be explored. Thus, the comparison between self-assessment and summative exams becomes a tool to detect test anxiety or overestimation and necessary measures can be initiated.

III. RESULTS

A central part of our framework of formative assessment is the diagnostic tool. In order to detect concepts and misconceptions it must be properly designed. Here we applied the 1992-version of the force concept inventory (FCI) [8] to two classes of the high school Romanshorn in Switzerland. Although the FCI has been criticized [9] it is still the most valid and reliable tool to assess students understanding of concepts in mechanics [10, 11].

The two classes are named class X and class Y. They should not be confused with the numbered classes below from latent class analysis who are subgroups of class X or class Y. At the time point of the FCI administration class X was right before taking an introductory, algebra-based mechanics course. In contrast, for class Y the same course has been finished two months ago. Both classes have a size of 18 students. Class Y was an immersion class (physics is taught in English) where the original FCI was applied whereas class X is a German class where a translation was used.

The answers of the students were coded in 0,1 matrix (false answer: 0; correct answer: 1) and grouped into concepts following the partitioning of Hestenes *et al.* [8]. First we analyze class Y (after the course) with an average of 16.3 points out of 29 questions. In our opinion, a diagnostic tool should have an easy read out for the students. We also explored Facet Innovations, which provides an interesting diagnostic tool for an introductory mechanics course [12], however the read out was much too complicated for the students. Thus, we suggest that the output of a diagnostic test should have the form shown in Fig. 3.
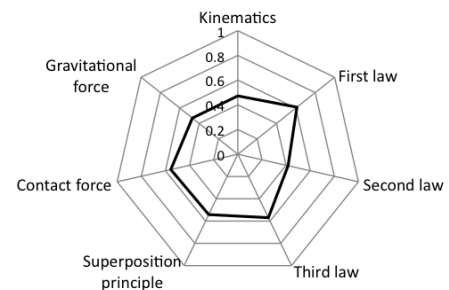
A
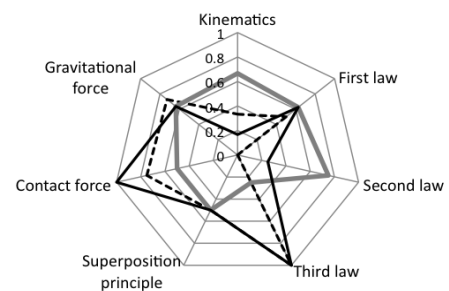
B

FIGURE 3. Suggested feedback of the diagnostic tool for the students (A) and for the teacher (B). Understanding of concepts for three students with a score of 17 (A), and the average performance of the class (B).

In Fig. 3A the average percentages of correct answers of three students of class Y for the seven concepts are displayed. They all had the same score of 17 points, which almost corresponds to the average performance of the class. Despite the identical score of the three students the individual profiles are quite different. The student associated with the dashed line got all questions concerning the third Newtonian law right, however, he completely missed Newton's second law. For the student who is represented by the grey line it is almost opposite.

The response of the whole class can be seen in Fig. 3B. This is the diagram serving as feedback for the teacher. If the FCI is considered as a diagnostic test it should be followed by a reflective lesson. From the graph 3B the teacher can read out how to prepare and organize this lesson.

Due to the fact that the data of clicker questions, as well as those of the diagnostic tool (including FCI) are saved electronically we can use this data to follow the students' individual formation of concepts or of conceptual change in a time dependent fashion. The goal of the study is to get a better understanding on how these concepts are acquired and how the students deal with the simultaneous presence of concepts and misconceptions. Moreover we would like to use the longitudinal data to identify classes with the same behavior. If indeed different classes can be found then specially designed programs have to be developed for those classes.

In our preliminary study we again used the data of class X and class Y to detect classes with the same degree of understanding in the data set. Therefore, we have applied latent class analysis (LCA) to the class Y that solved the FCI after the physics course. The R statistics environment provides a program for LCA [13]. A reasonable fit was found for four classes using the Akaike information criterion.
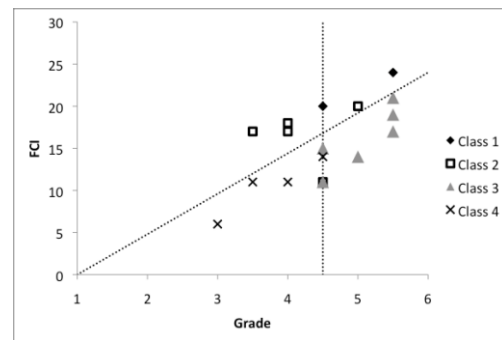
A
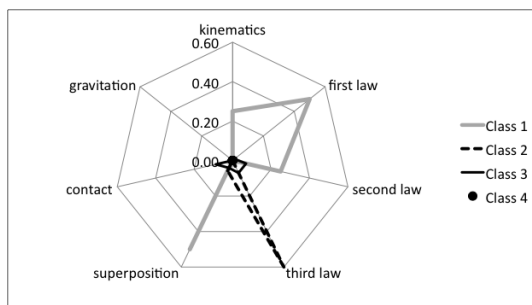
B

FIGURE 4. Latent class analysis (LCA) of the result of the FCI administered to class Y. Class 1, class 2, class 3 and class 4 represent, respectively, "problem solvers", "attentive students", "opportunists" and "conceptless students". The graph in (A) shows the probability that all problems related to the same concept are solved correctly. (B) Location of the classes in the graph FCI-score versus semester grade. The dotted lines show the division of the plane into four domains representing the four classes.

The output of LCA lists for each class the item response probabilities of the problems. If a concept is properly grasped a series of problems referring to the same concept have to be solved consistently. Thus for a problem set associated to a concept the response probabilities have to be multiplied. The results for the four classes are shown in Fig. 4A. The different classes that appeared in the analysis can be assigned to four groups of students typically present in a class with a traditional teaching style. There are the "problem solvers" (class 1: These students understand all concepts, which are relevant to solve problems and neglect all others), the "attentive student but poor problem solver" (class 2: Students that try to get all concepts, but have severe difficulties with concepts used in problem solving), the "opportunist" (class 3: Students who have for all problems the chance to solve it right; they have the best grade to knowledge ratio) and the "conceptless student" (class 4: students who are not able to apply a concept in different situations properly).

Of course, these are preliminary results, which should be treated with caution. However, also in the graph (Fig. 4B) FCI-score vs. grade (lowest grade 1; highest grade 6; 4 is the boundary between sufficient and insufficient) it seem that these four classes appear. The classes can be associated to the four quarters of the graph (class 1: Upper right; class 2: Upper left; class 3: Lower right and class 4: Lower left). This distribution occurs since the teaching style applied in Switzerland still fosters problem solving capabilities rather than conceptual understanding. Therefore the best students (concerning grades) are in class 1 and class 3. Due to the small data set we accept the two outliers of class 2.

Of particular interest is the analysis of misconceptions of a class before taking the physics course. To detect the spectrum of misconceptions we used the results of the FCI of class X (average of correct responses is 7). The responses are transferred to the 79 possible answers related to misconceptions ([8], Table II) and again coded in a 0,1

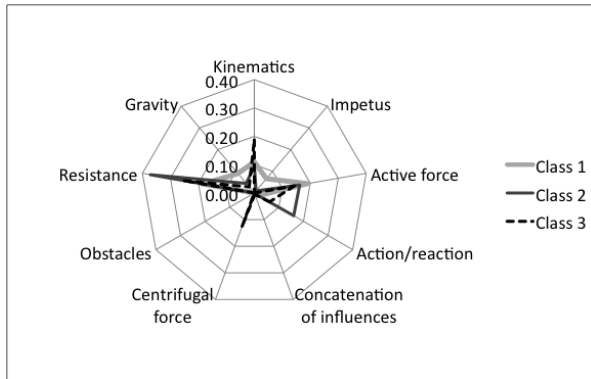matrix. The new data set is then fed to LCA allowing the program to search for three different classes.

FIGURE 5. Latent class analysis (LCA) of the result of the FCI administered to class X (no high school physics course). Three classes of misconception profiles were detected class 1: 3 students, class 2: 11 students and class 3: 4 students.

In order to detect solid misconceptions, where several problems have a high probability to be answered in a consistent fashion, we multiplied the probabilities of the LCA output for a given misconception subcategory. For example, if we take the subcategory I1 (category: impetus; subcategory: impetus supplied by a hit) the probabilities to answer question 9 by B or C, question 22 by B, C or E and question 29 by D are multiplied. If the multiplication provides a number close to 1 we would speak of a solid misconception. In order to represent the data we averaged over categories (coefficient of category impetus: average of I1 to I5). The results are shown in Fig. 5. The number of students in class 1 class 2 and class 3 are 3, 11 and 4, respectively. Due to the small sample size we are not able to give a detailed analysis of the model selection.

The results suggest that class 1 has misconceptions in the categories kinematics, impetus, active force, resistance and gravity. The misconceptions for class 2 and 3 are almost equal (kinematics, active force, action/reaction and resistance) with the exception of the centrifugal force misconception of class 3. The analysis suggests that classroom activities with an emphasis on the common misconceptions might be beneficial.

Since the students took already a natural science course in the middle school (K-8) the analysis might also lead to improve teaching in the educational institution visited before.

VI. CONCLUSIONS

We have developed a framework of formative assessment in high school teaching of physics. It consists of four branches: (1) Monitoring students' learning by self-assessment, (2) formative classroom activities to induce

classroom discussions and peer instructions, (3) diagnostic tests and (4) the formative use of summative assessments. Classroom activities have to be designed in a way that the learning of concepts is fostered. This can be accomplished for example with the clicker system. Diagnostic tools are important reference points to detect learning deficits. Reflective lessons following the administration of a diagnostic tool have to be designed to close the gap between the actual knowledge and the learning goals. Finally summative tests are used in a formative way to get an additional feedback about problem solving strategies and about the correct application of self-assessment. We also think that formative assessment tools dynamize teaching. It increases student activities during lesson, fosters classroom discussions and peer instructions. The results of formative assessments and diagnostic tools can also be used to follow students' acquisition of concepts and the change of concepts in time.

In our preliminary study we applied the FCI to two classes of the High School Romanshorn, Switzerland, one before (Class X) and one after the physics course in mechanics (Class Y). Latent class analysis for class X revealed the presence of solid misconceptions of the majority of students. Although we have identified three different classes there is a large overlap of common misconceptions. They are in the categories kinematics, active force, action/reaction and resistance. Teaching has to take these misconceptions into account by designing classroom activities where students are given the opportunity to remedy their ideas.

Analyzing class Y by LCA in conception space showed the appearance of four classes: "problem solvers", "attentive students", "opportunists" and "conceptless students". We think that the occurrence of these four classes is the generic outcome of a teaching style which fosters problem solving capabilities rather than conceptual understanding. If this is indeed the case will be further investigated in a large-scale experiment.

REFERENCES

[1] Black, P. and Wiliam, D., *Assessment and Classroom Learning,* Assessment in Education **5**, 7-71 (1998).
[2] Dufresne, R. J. and Gerace, W. J., *Assessing-to-learn: Formative assessment in physics instruction,* The Physics Teacher **42**, 428-433 (2004).
[3] Schneider, M. and Stern, E., in *The Nature of Learning*, edited by Dumont, H., Istance, D. and Benavides, F., (OECD, 2010).
[4] Mazur, E., *Fareweel Lecture?,* Science **323**, 50-51 (2009).
[5] Black, P. and Wiliam, D., *In praise of educational research': Formative assessment,* British Educational Research Journal **29**, 623-637 (2003).
[6] Keeley, P. D., *Science Formative Assessment* Corwin Press, (NSTA Press, Thousand Oaks, CA, US, 2008).

Clemens Wagner and Andreas Vaterlaus

[7] Threlfall, J., *The formative use of assessment information in planning–the notion of contingent planning,* British Journal of Educational Studies **53**, 54-65 (2005).
[8] Hestenes, D., Wells, M. and Swackhammer, G., *Force concept inventory,* The Physics Teacher **30**, 141-158 (1992).
[9] Huffman, D. and Heller, P., *What does the force concept inventory actually measure?,* The Physics Teacher **33**, 138-143 (1995).
[10] Hake, R. R., *Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,* American Journal of Physics **66**, 64-74 (1998).
[11] Stewart, J., Griffin, H. and Stewart, G., *Context sensitivity in the force concept inventory,* Phys. Rev. Spec. Top.-Phys. Educ. R. **3**, 010102 (2007).
[12] Minstrell, J., Anderson, R. and Minstrell, J. E., Facet Innovations http://www.diagnoser.com/diagnoser/ (2008).
[13] Chung, H., Retrieved on 7/6/2011 from https://www.msu.edu/~chunghw/downloads.html (2007).