www.sails-project.eu

Final report on the assessment frameworks and instruments for IBSE skills

SAILS
Strategies for Assessment of Inquiry Learning in Science

Strategies for Assessment of Inquiry Learning in Science

D2.4 Final report on the assessment frameworks and instruments for IBSE skills

Authors: Csaba Csíkos, Benő Csapó, Gábor Veres, Ferencné Adorján, Katalin Radnóti, Christine Harrison, Brian Matthews, Paul Black, Eilish McLoughlin, Paul van Kampen, Deirdre McCabe, Odilla Finlayson

Project name: Strategies for the Assessment of Inquiry Learning in Science (SAILS)
Project number: 289085
Start date: 01/01/2012
Duration: 48 months
Lead partner for this deliverable: University of Szeged
Project coordinator: Dublin City University
Contact: info@sails-project.eu
Website: www.sails-project.eu

The research leading to these results has received funding from the European Union's Seventh Framework Programme for research technological development and demonstration under grant agreement no 289085.

Contents

Introduction

A key aim of the SAILS project is to present a framework for assessment of inquiry learning in science. The purpose of this framework is to provide a detailed description of the content of assessment and to describe what and how to assess in the context of IBSE. In this work the frameworks may be used as models, in particular as a way of presenting assessment content and examples for the assessment items.

In an earlier deliverable of this work package (D2.1) a taxonomy of objectives to be assessed was provided. In the report D2.2, the structure of the proposed framework for SAILS was outlined. The structure was operationalised in order to develop the initial assessment items for piloting and trialling. Three topics were selected that provided teachers with a view of the type of assessment opportunities that could be developed within each of the subject areas (physics, chemistry and biology) and provided specific examples of what inquiry skills to assess and how to assess them . The topics chosen for the initial assessment items were Food, Rates of Reaction and Speed aimed at the lower second level students as they are applicable across most countries. The initial assessment items were presented in D2.2 and a report on their trialling with teachers has been reported in D3.1.

In the previous report D2.3, the assessment practices and instruments of four highlighted inquiry skills as well as further dimensions of the cognitive learning outcomes were illustrated by means of selecting examples of practice from different countries' reports, showcasing a range of assessment approaches. These experiences from across several different countries have been collected using unified report templates thus allowing for syntheses of the approaches adopted (called SAILS units). This strategy provided the project the opportunity to draw conclusions from cross-country experiences, and to showcase strategies for assessment of inquiry learning in science across Europe. From this analysis, it was clear that we had to develop a common understanding of particular inquiry skills within the consortium and also to show in more detail the strategies used in assessing particular skills. Therefore, several draft units were developed by members of the consortium, trialled by different partners, and reported on in the form of case studies. Following further evaluation and testing with teachers, these draft units and case studies are combined into SAILS units that give details of inquiry lessons with embedded assessment modes and criteria. These experiences provide the illustrative examples that inform this framework document.

D3.3 provides detailed information about the development of the SAILS units (see Figure 1.1 in D3.3) which is the information source for the current D2.4 as well. It means that when citing examples of different assessment practices and instruments it can now be done by selecting teaching unit that has been trialled in at least three different SAILS-countries. Consequently, the main novelty (and in our belief, strength) of the current document lies in the illustrations provided for each cognitive learning outcome in the framework.

The assessment framework generally answers two main questions: what to assess and how to assess? Therefore Section 1 of this document discusses Inquiry Learning in science while Section 2 discusses assessment and assessment strategies.

A high-level outline of the framework SAILS has adopted to assess inquiry skills is depicted in Figure 1. It illustrates the interplay between the inquiry skills and competences targeted in both the rational/logical and social/emotional domains (described in Section 1), suitable assessment practices for these (described in Section 2), the assessment strategies used (described in Section 3) and the SAILS inquiry and assessment units, examples of which are given in Section 4.

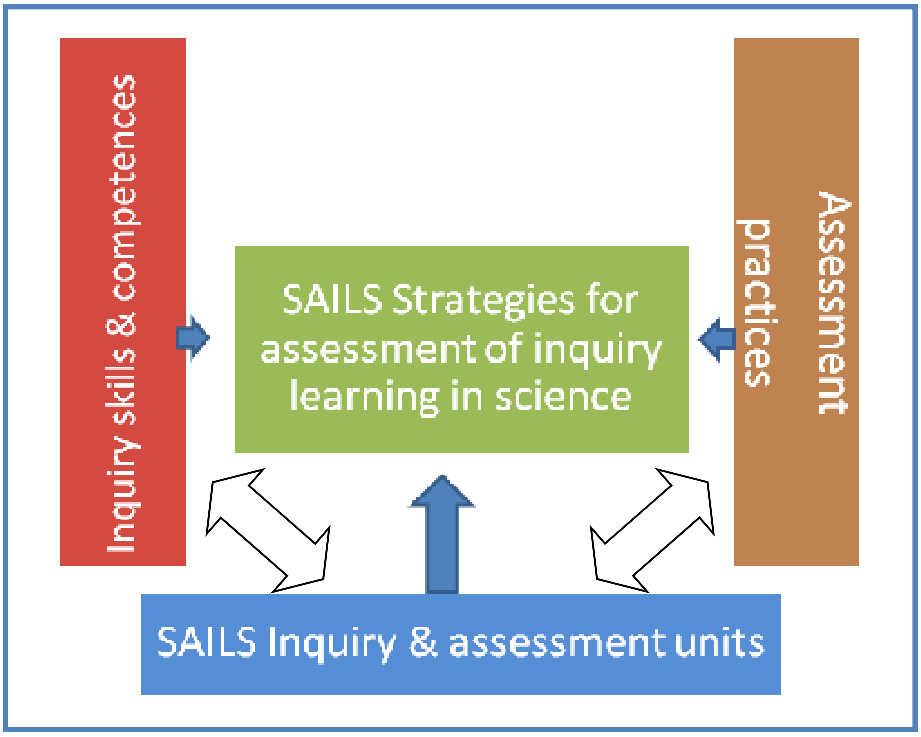

Figure 1 SAILS framework for assessment of inquiry learning in science

1. Inquiry Learning in Science

Inquiry based science education (IBSE) is an approach to teaching and learning science that is conducted through the process of inquiry. Some of the key characteristics of inquiry based learning are (Kahn & O'Rourke, 2005):

- Students are engaged with a difficult problem or situation that is open-ended to such a degree that a variety of solutions or responses are conceivable;
- Students have control over the direction of the inquiry and the methods or approaches that are taken;
- Students are engaged in dialogue with others to explore approaches and decisions about the inquiry;
- Students draw upon their existing knowledge and they identify what their learning needs are;
- The different tasks stimulate curiosity in the students, which encourages them to continue to search for new data or evidence;
- The students are responsible for the analysis of the evidence and for presenting evidence in an appropriate manner which defends their solution to the initial problem.

Within an inquiry culture student learning is especially valued; students are fully involved in the active learning process. Students who are making observations, collecting data, analysing data, synthesising information, and drawing conclusions are developing problem-solving skills. These skills fully incorporate the basic and integrated science process skills necessary in scientific inquiry. The students are also involved in working collaboratively, learning how to deal with differing opinions and perspectives and becoming self- and co-dependent. Through inquiry learning students develop the lifelong skills critical to thinking creatively, as they learn how to solve problems using logic and reasoning. These skills are essential for drawing sound conclusions from experimental findings.

While there are a range of skills and competencies developed through inquiry, within the context of the current document, the focus is on how content knowledge, inquiry skills, reasoning ability and scientific literacy may be assessed within an inquiry lesson.

This model, based on cognitive research and taking into account the experiences of framework developmental processes of PISA and TIMSS (see Csapó, 2010) has been applied earlier in the framework development in mathematics (Csíkos & Csapó, 2011) and science (Csapó, 2012; Korom, Németh, Nagy & Csapó, 2012). This model has been adapted end extended for the assessment framework presented here.

One dimension of scientific knowledge to be acquired by students while learning in the context of IBSE is, of course, the scientific concepts and other content elements prescribed for them in the curriculum. Test theory and instructional practice have several decade old traditions in assessing the quality and amount of factual knowledge to be learnt in school. The amount of factual knowledge seemed to be crucial in previous decades but now the development of and the quality of students' skills and knowledge are of even bigger importance than before.

A second dimension refers to cognitive processes that contain more or less automated thinking skills. In each school subject and domain, a bunch of cognitive skills and abilities are needed to be performed. Traditionally, the term skill refers to more or less routine thinking processes; the term ability refers to any thinking and reasoning processes that have measurable individual differences. In the literature (see e.g., Overton, 1990), thinking is described as a more general term, while reasoning stands for thinking processes that involve inferences. There are several reasoning skills and abilities that are obviously needed and applied in science classes: inductive reasoning, deductive reasoning, combinatorial and probabilistic thinking should function smoothly in order to make sound inferences independently of the actual content of a lesson.

Any type of knowledge should be applied in different contexts. The PISA studies defined several types of situations or contexts in which knowledge should be applied. For instance, when official documents or announcements are used in a scientific task, the task evokes a "public" situation or context. It has been widely documented that knowledge transfer from a given situation to another is far from being easy or obvious. Consequently, the term literacy has been closely associated with knowledge application in different contexts (Csíkos & Verschaffel, 2011).

Inquiry skills, albeit discussed in separate sections, overlap with other knowledge components described in the framework. For example, scientific concepts play their role in each phase of scientific inquiry, thus contributing to and taking part in each inquiry skill. Reasoning abilities are necessary for making sound and coherent arguments while working either individually or collaboratively. Some of the tasks trialled in several partner countries have detailed connections to out-of-school contexts. Both the aim of fostering the application of knowledge and skills in different situations and the aim of facilitating students' participation, activity, enthusiasm and motivation (affective learning outcomes) resulted in tasks that addressed a variety of contexts, thus possibly improving the transfer of students' knowledge from one context to another. So what is the process of IBSE and what are the inquiry skills and competencies developed?

1.1 Curricular content

Before discussing inquiry skills, we briefly consider the issue of curricular content. Inquiry-based science education must fulfil the criteria of providing appropriate scientific knowledge. As the participating countries of the SAILS project differ in school structure, curricula requirements and assessment practices (see Deliverable 1.2), we focus here on some common and important scientific concepts, and how the assessment of conceptual understanding was addressed in different case studies in lessons.

There has been a world-wide shift from traditional teacher-centred pedagogical approaches to student-centred active learning. This results in changes in the teaching of scientific concepts, for example through inquiry processes. One particular approach considers two phases of the science class of particular importance in developing students' conceptual understanding. Firstly, when approaching a problem or planning an experiment at the beginning of the lesson, the necessary prerequisite knowledge components may be evoked through discussions/brainstorming activities. Students are usually allowed and encouraged to freely use any terms they think to be related to the current problem. This brainstorming or mind mapping collection of ideas often reveals students' misconceptions. Commonly used scientific terms such as speed, heat, impetus, or life have a rich semantic framework developed from students' past experiences and former studies. Secondly, having conducted active observations and experiments, students reflect on their findings, often in groups, using appropriate scientific terms. In many cases, due to the experiences gathered, students themselves feel the need to use more precise terms at this stage of the learning process.

1.2 Inquiry skills

In the SAILS project, the following definition given by Linn and Davis (2004) is followed. This definition has been further discussed and elaborated in a WP1 milestone draft of the SAILS project (Draft report on key skills and competencies).

> *Inquiry is the intentional process of diagnosing problems, critiquing experiments, and distinguishing alternatives, planning investigations, researching conjectures, searching for information, constructing models, debating with peers, and forming coherent arguments.*

The *genus proximum* of this definition is "intentional process", while the *differentia specificae* are the inquiry skills. Firstly, the use of the term *intentional* refers to conscious, strategic processes that may and should be generalizable throughout different domains and contexts. Secondly, the term *skill* can be reliably used since mastering of these processes are different from knowing merely when or how to use them. Several authors, (e.g. Fradd et al (2001), Lee et al (2001), Sutman & Saxton (2001), and Wenning (2007)) suggested well-defined classifications of inquiry skills, while Bybee (2009) structured an inquiry process into inquiry phases, named: engagement, exploration, explanation, elaboration, evaluation (5E model).

From the Fradd et al. (2001) and Wenning papers (2007), a comparative table can be derived (Table 1). This table simplifies both taxonomies in order to make them comparable. The main message of this table is that taxonomies of inquiry skills may be anchored to the consecutive phases of scientific inquiry.

From these lists (Table 1) and Linn and Davis (2004), the four inquiry skills were chosen with the purpose of illustrating assessment strategies. The four inquiry skills addressed in this report are: planning investigations, developing hypotheses, working collaboratively, and forming coherent arguments.

These were chosen for a number of reasons. The rational skills, such as planning and hypothesising, are emphasised in many curricula, but there have been many criticisms of this approach and it is accepted that inquiry is much more than a logical process (Barton, 1998; Baumfield, 2006; Bernstein, 1983; Krajcik, Blumenfeld, Marx, & Soloway, 2000) and include aspects such as collaboration, communication and critical thinking. So, in line with Linn and Davis, the four inquiry skills chosen are a combination of both rational and social.

TABLE 1 A COMPARISON OF TWO WIDELY RECOGNIZED TAXONOMIES OF INQUIRY SKILLS.

“Wenning-skills”	**“Fradd-skills”**
Identify a problem to be investigated.	Questioning
Formulate a hypothesis.	
Design experimental procedures to test the prediction.	Planning
Conduct a scientific experiment; collect meaningful data, organize, and analyze data accurately and precisely.	Implementing
Apply numerical and statistical methods to numerical data to reach and support conclusions.	Concluding
Using available technology, report, display, and defend the results of an investigation to audiences that might include professionals and technical experts.	Reporting
	Applying

This framework outlines strategies that can be used to assess these skills and that can provide teachers with information that will help them to provide feedback and guidance to their students. Each skill is now discussed to highlight the SAILS understanding of the skill.

Firstly, these skills are often addressed and assessed within the case studies developed through the trialling of draft units by experienced inquiry teachers in each of the SAILS partner countries. Secondly, these four skills can be considered as representative of the different clusters of inquiry skills provided in the above-mentioned papers. A third perspective of focusing on some skills comes from the frameworks of international educational assessment surveys.

The SAILS inquiry skills include *Working Collaboratively* and *Forming Coherent Arguments*. These involve pupils in discussions with each other and so have to use social and emotional skills. However, collaboration with others can occur in all of the skills areas. For example, pupils can discuss how to plan or develop an hypothesis together. Working collaboratively involves a large range of skills that can be affected by a range of factors, such as personality and social class, and in particular gender. The gender dynamic formed part of the SAILS inquiry.

Inquiry in the classroom can take on several forms. One important dimension that characterises inquiry is the degree of autonomy students have. In SAILS, we have adopted the terminology of Walker (2007), see also 4.2.2. Thus, *guided inquiry* for activities in which activities are mostly prescribed by the teacher. Thus students have little procedural autonomy and may be unlikely to develop the skill of planning investigations, but may develop all other inquiry skills, such as forming coherent arguments – cookbook experiments are far from guided inquiry. In *open inquiry* students may formulate their own research questions, and plan and carry out experiments to answer them. *Bounded inquiry* lies in between these, often concerning activities where teachers provide a research question but students may plan and carry out the experiment. The boundaries between these three forms of inquiry are unsurprisingly open to interpretation. Generally speaking, the more open the inquiry level, the more likely it is that students develop the skills of developing hypotheses and planning experiments.

1.2.1 Developing hypotheses

Theoretically, developing hypotheses may be an inquiry phase that precedes planning investigations. In Fradd et al. (2001), questioning as an inquiry skill comes first and consists of two things: posing questions, and making hypotheses. Wenning (2007) drew our attention to the complementarity of thinking processes that underlie this skill: inductive reasoning is used when formulating hypotheses, and deductive reasoning enables for making predictions from the hypotheses.

Formulating research questions often implicitly involves formulating hypotheses. Developing hypothesis involves extending a prediction to include a reasoned explanation/justification which is based on prior knowledge. The hypothesis should involve a testable question. Students, individually or in groups, should learn to formulate research questions that concern either comparisons between quantities or connections between variables. These two types of research questions can be interchangeably formulated in classroom discussion

There is no need to make a distinction between statistical and scientific hypotheses in the secondary school, since mere descriptive statistical analysis will surely suffice in almost all cases. However, when questions arise about the number of necessary repetitions (or sample sizes) teachers must be prepared to give appropriate age-related answers beyond the well-known motto: “one measure is never enough”.

1.2.2 Planning and implementing investigations

This skill (called “designing experimental procedures” by Wenning, 2007) refers to the intentional thinking processes necessary before beginning an experiment. Planning investigations can involve the following:

(a) Refining an open question so that it can lead to one or more specific questions;
(b) Deciding what you and others want to do to find out the answer to the question, including identification of variables and consideration of fair testing;
(c) Deciding what materials your group needs;
(d) Deciding how to record the information;
(e) Modifying the inquiry questions in response to ideas arising during the inquiry;
(f) In the light of final report, reflecting on what has been learned about the inquiry process during planning.

These steps are not linear, questions and plans are adapted and refined through the inquiry. As information is collected and observations made, it becomes clear that the plan is inadequate, and it is then important to go back through the steps.

The (a) component of this skill points to the general question of how open the inquiry process should be. Fradd et al. (2001) showed that questioning is seldom left as the students’ responsibility; therefore students typically react to the question posed by their teacher. This usual sequence of the science classes justifies that the assessment of planning investigations usually comes first, and the other inquiry skills are assessed later. Why we emphasize this chronological order is that in “real” scientific research design (see Kirk, 1995) testable formulations of scientific hypotheses precede the specifications of the experimental design.

The components of this inquiry skill presume an appropriate level of reasoning skills. For instance, when making decisions on what to change and what to keep constant in an experiment, combinatorial thinking plays its important role.

1.2.3 Working collaboratively

Observing and assessing the quality of collaborative work is still a great challenge in educational research. At the classroom level, however, it is possible and desirable to assess the quality of both individual and collaborative efforts. Collaborative learning involves students sharing ideas and using their peers as a resource. Collaboration implies that individual perspectives are valued and considered but not necessarily lost as the group idea develops. Rather it infers that ideas are evaluated, compared and sometimes reshaped and expanded to form new ways of thinking or to reach decisions with greater confidence. To be effective in a collaborative learning environment, learners need to demonstrate a growing development in their listening skills, the ways they express ideas and how they interact socially and emotionally with others. In all cases these social and emotional interactions are overlaid with gender and/or social class, and/or personality differences and are affected by cultural beliefs and values.

It should be emphasised that collaborative learning is much more than allowing students to work on an inquiry in groups. The range and type of skillsets used within a collaborative learning environment enable learners to contribute towards the development of scientific argument, a consideration of the strength of evidence and the communication of the impact and implications of inquiry findings and these are all skills which contribute to life skills that are appropriate for 21st century scientific thinking.

The skills required for working collaboratively are slow to develop and require that students are provided with constructive guidance and feedback at relevant times within this development. Within these it is important that the natures of the interactions, such as gender, are made explicit and open to debate and so change (Matthews, 2006). Therefore, it is important that teachers create a range of opportunities for students to work collaboratively and to foster a system of feedback that responds to student needs and progress. Within a single lesson it is often inappropriate for teachers to try to assess or feedback to each and every individual student or the whole class, rather teachers may focus on specific individuals or groups. In subsequent lessons the teacher may feel that the focus needs to be on different specific students or groups. In this way, the teacher is able to provide guidance to all students when they need it most.

1.2.4 Forming coherent arguments

This inquiry skill builds upon the domain-general reasoning processes described in Section 1.3 of this report. Coherent is an important word, therefore arguments need to include all evidence (e.g. explain outliers). There needs to be critical discussions on the coherence of the arguments developed. An important part of this skill involves sorting out of previous knowledge, determining relevance and making relationships. Forming coherent arguments can be developed at all stages of an inquiry (e.g. planning an investigation, developing hypothesis and explaining results). There are two reasons for explicitly defining this skill. Firstly, as an inquiry skill, forming coherent arguments is a content- and context-bound intentional process. The quality of argumentation depends on the characteristics of the task (e.g. the presence or absence of prerequisite knowledge, and the cognitive load demand). Secondly, similar to the previous skill, forming coherent arguments (or the lack of it) can be observed and assessed throughout the implementation, analysis and reporting stages of inquiry. Students should develop skills to form coherent arguments individually and as a group. The group interactions can play a part in the decisions made, for example, if there is a particularly dominant personality. Similarly, it has been found that girls' voices can be heard less than boys and boys' utterances are given more value (Thomas & Wareing, 1999). Hence teachers can help pupils develop the skills required for argumentation through making explicit the interactions, and then getting pupils to respond in ways that make their interactions better.

A further important component of this inquiry skill is the capacity to decide which kinds of evidence are supportive (verifying) or falsifying. To make distinctions between verifying, falsifying or nonaligned evidences requires critical thinking, i.e. reflective thinking about why we accept or refuse a statement or a conclusion (Norris & Ennis, 1989; Ennis, 1995; Aktamış & Yenice, 2010; Eklöf, 2013).

1.3 Reasoning skills and abilities

Scientific reasoning is often referred to as the most advanced form of human thinking. In Deliverable D2.1, “Report on the Strategy for the assessment of skills and competencies suitable for IBSE”, reasoning skills necessary for scientific inquiry were identified and described, e.g., deductive (logical) reasoning, inductive reasoning, combinatorial and probabilistic reasoning. Scientific reasoning uses abstractions and symbols, and represents phenomena in variables and dimensions. Scientific reasoning analyses the relations between the identified symbols and variables, and in these analyses, reasoning skills described in previous sections are applied. For example, scientific reasoning often deals with ratios, proportions and probabilities; designing experiments requires systematic combination of variables involved.

Scientific reasoning is often manifested through dialogue which involves reframing one’s own thinking through the process of argumentation. Argumentation requires the dynamic organization and re-organization of evidence, data, figures, etc., to persuade and convince one’s self and others. Argumentation is a core aspect of the nature of science. Peer-review and exchange is integral to the development of scientific ideas. The dynamics of these interactions may be influenced by factors, such as culture, gender and existing relationships. Scientific reasoning involves carrying out logical operations and establishing causal relationships between observed changes and may include inductive or deductive reasoning.

In classroom settings, the “reasoning errors” of thinking may be observed and mediated. For example, when separating variables of an experiment, keeping one or more variables constant while changing others, combinatorial reasoning is crucial. In cases where two variables are to be manipulated, understanding that there are at least four possible situations is necessary. Even when students have previously encountered combinatorial thinking, i.e. listing and identifying all cases, they can struggle to translate it to different contexts.

1.4 Scientific literacy

The literacy concepts used in practice/everyday life are unique both in terms of employing and interpreting different concepts and defining objectives. Nevertheless, the scientific literacy frameworks and standards bearing different objectives and relying on the traditions of a particular culture and education system exhibit several similar features. For example, scientific literacy is commonly considered to entail much more than the integration of knowledge, values and the fundamental elements of scientific education, as it is a complex and multi-dimensional knowledge structure (Roberts, 2007). There is a broad consensus that scientific literacy is science knowledge that has a bearing with regard to both individual and social aspects. PISA studies (PISA, 2015) have highlighted the importance of thinking processes that go beyond the mere recall of factual knowledge and the immediate use of routine algorithms. Individuals need transferable, expandable and adaptive science knowledge/scientific literacy. Scientific literacy incorporates the idea of individual and societal usefulness of knowledge, therefore the knowledge components learnt in school should be transferrable to new contexts and for several different purposes. Scientific literacy in PISA 2015 is defined by the three competencies to:

- Explain phenomena scientifically (recognize, offer and evaluate explanations for a range of natural and technological phenomena);

- Evaluate and design scientific enquiry (describe and appraise scientific investigations and propose ways of addressing questions scientifically); and
- Interpret data and evidence scientifically (analyse and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions) (PISA, 2015, P.5-7).

This definition tends to reinforce the traditional rationalist view of science, rather than placing the multi-dimensional and social aspects at the centre. However, the use of IBSE can help students to develop most of the skills of scientific literacy. Scientific literacy tasks are connected to relatively long, real-life texts, requiring students to distinguish important and distracting data, actively use their prior factual knowledge and reasoning skills developed through both formal and informal learning. Scientific literacy also requires people to debate ideas (Forming Coherence Arguments) and their connection to society; it encompasses contributions to a democratic dialogue across diversity (Working Collaboratively). Scientific literacy is particularly important within an inquiry class when finding problems and posing relevant questions, and when considering the generalizability of the findings in an experiment. Of course, in other inquiry phases, e.g. in planning an investigation, knowledge transfer occurs as a result of finding analogies between real-life experiences and the constraints of the current problems, the literacy components of scientific knowledge are utilised.

2. Assessment in the context of the SAILS project

Previous deliverable documents (particularly D2.1 and D3.1) have given an extensive overview of assessment. This section provides a brief overview of the practices of assessment as interpreted by the SAILS project.

Educational assessment is a well-defined field of research and practice which deals with collecting, analysing and utilizing data on students' learning outcomes (Black, 2000). A great variety of methods and instruments are available for educational assessment and measurement, and most of these are appropriate in the context of IBSE as well.

Discussions of assessment usually distinguish between the two main purposes of the assessment; namely formative and summative. The characteristics mentioned by Harlen and Deakin Crick (2002) appear in most descriptions:

> "Assessment is a term that covers any activity in which evidence of learning is collected in a planned and systematic way, and is used to make a judgment about learning. If the purpose is to help in decisions about how to advance learning and the judgement is about the next steps in learning and how to take them, then the assessment is formative in function. If the purpose is to summarize the learning that had taken place in order to grade, certificate or record progress, then the assessment is summative in function." (Harlen & Deakin Crick, 2002 p. 1).

Assessment of IBSE skills and competencies requires teachers to be able to use a variety of tools to determine where students are in their learning. From these data, they can make judgements that can help the student to decide on the next step in learning, and so guide them towards improvement. This data can often be used in both formative ways and summative ways – the formative to provide feedback and guidance for future learning and decisions about teaching; the summative to enable teachers to chart progress of inquiry skill development over time.
When interpreting assessment in the context of IBSE, we have to take into account that inquiry methods consider learning as an active and constructive process, thus assessment feedback should be integrated into this active learning process. Thus, the assessment has to deal with components of

knowledge and skills where changes are observable after relatively short periods. In such situations, there may be more direct correspondence between classroom activities and learning outcomes, and the feedback the assessment provides may orient the next phase of learning. However, skills such as working collaborative and teamwork require time to develop, especially with regard to gender and other cultural issues.

In a cyclical model of assessment (Figure 2), Harlen (2005) emphasizes the main function of formative assessment in step 3 and 4, namely, how it is used for learning, i.e., planning the next appropriate steps in the teaching and learning process. The cycle as shown by Harlen can be repeated many times within any teaching episode.

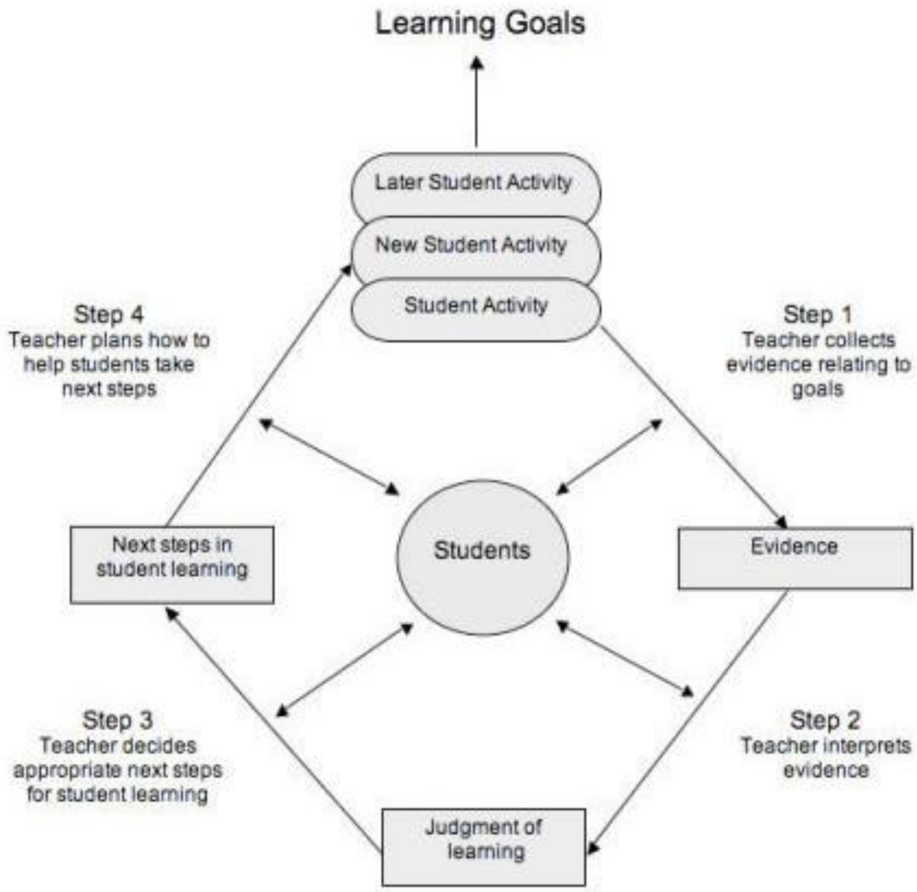

Figure 2 Assessment Cycle (Harlen, 2005)

While some aspects of assessment were dealt with in similar ways to what teachers normally do in science lessons, such as tests and marking lab reports, several of the teachers began to strengthen their classroom assessment practices. This resulted in a more student-centered approach to assessment, which engaged learners more actively in the assessment process. Some of the German teachers used green, orange and red cups, during the inquiry, which students used to signal to their teacher when they needed help (red cup), when they still working out ideas (orange cup) and when they were working easily towards their goal (green cup). By stacking the cups, with the signal cup on top, the teacher could quickly see which groups needed help and which were okay to continue without any intervention.

Other teachers worked on developing the classroom dialogue in lessons, so that teachers could probe understanding during the inquiry, as well as observe what different groups of students were doing. The UK group developed a range of questions to support teachers in carrying out this dynamic collection of assessment data during the inquiry cycle. This enabled them to both track how different groups were planning and implementing their inquiry and to make decisions on when they should stop the various groups to share ideas between the various groups in the class.

Over the course of the inquiry , the teachers were able to build up various layers of assessment data at various stages in the inquiry cycle– some data accessed from individuals and groups during the inquiry, some from the ways each group reported on their progress and intentions and then, finally, some data from their final reporting of their findings. In this way, the teachers built up rich pictures of how individuals and groups performed throughout the inquiry.

As an example of how teachers may use their knowledge of students' attainment, in Case Study 3 (CS3) of the *Polymers* unit the teacher selected the groups based on students' previous results and organised them so that students were in mixed ability groups. This enabled students to work more collaboratively, with higher attaining students supporting lower attaining students in developing inquiry ideas. At the same time, discussion with peers who were still tentative with their understanding helped the higher attaining students in articulating their more sophisticated ideas and so consolidated their understanding. This fits with Heritages's idea of students utilising one another as a resource (Heritage, 2007).

In a more recent paper, Heritage (2013) describes how data can be collected about student understanding. Interaction is identified as a primary source of evidence about understanding: first of all, the interactions between students and teacher, but further sources such as students' writings, drawings and other artefacts can also be analysed. It is from this rich data source derived from classroom activities that teachers can build a picture of a student's developing competence in inquiry and this will inform their choices in which activities to focus on next and which skills to emphasise through their teaching. For the learner this feedback directs them to where to focus their effort so that they are more likely to make improvements over the next phase of learning.

2.1 Assessment in the inquiry classroom

Assessment in the inquiry classroom can take many different forms. It may be useful to discuss these under headings of: What will be assessed, when is it assessed and by whom is the assessment carried out?

- **What will be assessed?**

Within the SAILS framework it is recognized that a wide variety of artefacts and activities may be assessed. One category comprises student artefacts: not just reports, but also mind maps (see e.g. CS1 of the *Electricity* Unit), the plans for an investigation (see e.g. the *Up There How is It?* Unit), drawings (see e.g. the *Light* Unit), models (see e.g. the *Natural Selection* Unit). However, activities such as dialogue (can students articulate their thoughts to one another, can groups negotiate a solution to conflicting ideas; see e.g. CS6 of the *Oranges* Unit) are also assessable and form part of the framework.

- **When is it assessed?**

Assessment may take place pre-activity – be it in the form of a pre-test, a brainstorm, a discussion; during an activity – for example; by observing student-student dialogue or by evaluating measurements or observations; or post-activity – not just by taking up reports and grading them, but also by students evaluating their own investigation (see e.g. the *Woodlice* Unit).

- **By whom is the assessment carried out?**

Throughout the SAILS units there are examples of assessment done by the teacher (almost every unit), but students may also evaluate each other's work (peer assessment, see e.g. the *Global Warming* Unit) or their own work (self-assessment, see e.g. the *Food and Food Labels* Unit).

- **Use of the assessment evidence**

The emphasis in many SAILS units is on integrating assessment. It is therefore critically important that the teacher does not merely observe or evaluate the quality of the work or the process; what the teacher does with the information obtained often determines the quality of the assessment. Almost all case studies detail how the teacher used the assessment to improve the classroom experience.

2.2 Social contexts in assessment

There are various school cultures and classroom settings around the world and also in the countries participating in SAILS in respect of teaching methods and approaches to assessment. The large-scale international assessment projects have directed the attention of decision-makers to the importance of assessment, and in many countries national assessment systems have been implemented. This process has increased the level of expertise in assessment among teachers as well. However, the large scale assessments provide system level feedback, and the related analyses tend to have little impact on everyday classroom practices. One of the reasons behind this limited transfer is that immediate classroom level assessment requires different methods and instruments or different employment of these instruments in the learning context.

One of the major differences between large-scale testing and classroom assessment is that classroom assessment is more personal, takes place in a social context, and involves interpersonal communication. Developing teachers' assessment competencies may be one of the main avenues to improve quality, as it is conceptualized in the SAILS project. The first people who can assess students' learning outcomes are the learners themselves. Students' judgment of their own performance may be rather biased, and, as untrained assessors, they may make errors in the assessment process. Despite these constrains, students' self-assessment may be potentially very useful and also important taking into account the requirements of life-long-learning: Students are expected to become independent learners being able to manage the entire learning process, including assessment. As IBSE offers opportunities for students' individual work, it involves numerous possibilities to develop self-regulated learning strategies and metacognition as well.
In IBSE projects, self-assessment may take place in various forms, e.g. students may report and evaluate their successes and difficulties in the inquiry processes

Collaboration and teamwork is a typical setting of activities in modern societies, and assessing the related competencies has received growing attention. For example, in PISA 2015, collaborative problem solving will be the innovative assessment domain. Science inquiries may be carried out individually or in groups, and in both cases there are a number of contexts where competencies needed for collaborative activities may be fostered, with attention to gender (Matthews 2006). Similarly, there are several opportunities where students may evaluate each-other's learning, and give useful feedback to their peers. Using peer assessment may improve important social skills, communication skills and collaborative skills (Topping, 2013).

3. Strategies for assessment of inquiry learning in science

The strategies for assessment of inquiry learning in science developed by the SAILS consortium were focussed on the assessment of four key inquiry skills: developing hypothesis; planning and implementing investigations; forming coherent arguments and working collaboratively; as well as the key competencies of scientific reasoning and scientific literacy.

3.1 SAILS strategies for assessment

The strategies adopted for each of the inquiry skills and competencies within the SAILS project are summarised in Table 2.

TABLE 2 SAILS STRATEGIES ADOPTED FOR EACH INQUIRY SKILL AND COMPETENCE

SAILS Skills and competences	SAILS Strategies for Assessment
Developing Hypothesis	• Allow the pupils to play with the equipment in order to identify what variables should be considered/tested • Develop an appreciation that hypotheses are reasoned explanations that *could* answer the research question • Identify what makes an hypothesis testable by investigation or research • Support verbal, written and other formats for communication through structuring the inquiry to ensure these forms are used at different points. • Hypothesis should be logical and use precise language but based on evidence and/or prior knowledge
Planning and implementing Investigations	• Facilitate student brainstorming and sharing ideas through using pair-share • Give students time for testing and trialling of possible investigations • To encourage students to clearly articulate their plans and what variables they will investigate get them to explain their plans to other pupils • Challenge students to consider fair testing and the repeatability, reproducibility and reliability of their results • Challenge students to consider what materials/equipment are needed through the use of open questions that ask them to explain what they are doing • Encourage variation in how data/information is recorded through suggesting written, chart and picture forms. • Encourage students to reflect on their planning afterwards
Forming coherent arguments	• Require students to clearly articulate their viewpoint to each other • Mandate that all claims (rebuttals) must be based on observable (empirical) evidence and cover all observable evidence • Claims include conjectures, conclusions, explanations, models, or an answer to a research question • Allow students time for time assessing, critiquing, justifying and defending their evidence

	• Support verbal, written and other formats for communication. • Emphasize the importance of critiquing other people's ideas and how to critique others ideas (listening, respecting, providing quality feedback). • Students' reflection on the quality of their arguments should be encouraged. They should decide on what they need to do next time to improve their arguments. Take on formative feedback
Working collaboratively	• Define a lesson around a task that requires a collective effort • Give roles to pupils and assess how well they do them • Listen to groups and note interactions such as who listens, who supports others • Get pupils to self-reflect through writing on pro forma or rubrics • Peers to review other group's written and verbal methods/ or other aspects of inquiry • Discuss with each other how well they collaborated verbally or using written materials • Exchange messengers between groups to listen to explanations and explain to others as part of assessing listening and speaking skills • Teacher and pupils to observe and note if pupils interact and communicate with respect across diversity • Recognising listening and speaking through having turn taking • Get pupils to discuss if they could accept criticism and empathise • Teacher focus on specific students or groups who need teacher attention most in a particular lesson • Allow students varying degrees of autonomy
Scientific reasoning	• Encourage students to probe, find information, and seek explanations • Support students verifying scientific assumptions • Support verbal, written and other formats for communication. • Distinguish the differences between deductive and inductive reasoning • Provide open inquiries so that there are opportunities for both confirmation and disconfirmation
Scientific literacy	• Allow time for students to discuss/share their science knowledge • Use a wide range of multimedia formats • Facilitate "science talk" through using brainstorming, concept cartoons and place-mats. • Support student collaboration • Promote understanding of the nature of scientific knowledge • Develop appreciation of the joint enterprises of science and technology and the interrelationship of these with each and with other aspects of society. • When pupils work in groups make explicit how the social and emotional skills they are developing relate to participating in a democratic society.

3.2 SAILS methodology

The approach adopted by the SAILS project was to develop SAILS Inquiry and Assessment Units that can be used by teacher educators, with both in-service teachers and pre-service teachers, and that can support teachers extending their own assessment practices. Initially 34 different topics were proposed by the consortium that each comprised of inquiry activities and assessment suggestions, many building on materials that had been developed through other EU projects, such as PRIMAS, ESTABLISH and S-TEAM. From these suggestions, 19 topics were selected for development as SAILS Inquiry and Assessment Units on the basis that the topics were across the disciplines of physics, chemistry and biology; different types of inquiry were addressed – guided, bounded and open; and that these topics were appropriate for students from lower to upper second level.

The objectives of each SAILS Inquiry and Assessment Unit were to highlight to teachers the benefits of an inquiry approach in classroom practice; exemplify how assessment strategies are integrated in an inquiry approach; and illustrate the variety of assessment opportunities/processes that are available. In particular, each SAILS Inquiry and Assessment Unit would provide clear examples for teachers of how different inquiry skills can be assessed, alongside content knowledge, scientific literacy and scientific reasoning and illustrate the benefits of various types of assessments. In addition, the units will share examples of classroom practice, how evidence of student learning can be collected and evaluated using a variety of methods, e.g. classroom dialogue, teacher observation, student's written or multimedia work.

Each of the 19 selected topics were developed as a Draft Unit (DU) and then were trialled with teachers from at least three different countries from across the 12 participating countries in the SAILS consortium. The feedback from the teachers was collected in the form of Case Study (CS) reports which provided an account of how the DU was implemented in their classroom, what teaching approach was adopted, what skills were assessed and what assessment strategies were used. An overview of each unit, the content, implementation strategy and the assessment strategy used is included in Appendix A.

The breakdown of each Unit is given in Table 3, which identifies the science discipline and educational level of the unit, as well as, the skills and competences assessed in the 19 Draft Units (DUs) and in each of the associated Case Studies (CSs), and the country the trialling was carried out in. Over 100 case study (CS) reports have been presented by in-service teachers from across the twelve participating countries in the SAILS project which provide rich accounts of the variations in the teaching approaches and assessment strategies adopted for the 19 DUs. The reasons for these adaptations have been explained by the teachers as being due to several factors; including the teachers own inquiry and assessment confidence and competencies, the student's abilities and educational levels and the national curriculum and assessment strategies.

TABLE 3 OVERVIEW OF 19 SAILS INQUIRY AND ASSESSMENT UNITS, IDENTIFYING SKILLS AND COMPETENCIES ASSESSED IN THE CASE STUDIES (CS)

Unit Title	Discipline	Level	DU/CS	Country	Planning and implementing investigations	Developing hypothesis	Forming coherent arguments	Working Collaboratively	Scientific reasoning	Scientific literacy
Acids, bases, salts	Chemistry	Lower	DU	Greece	■					■
			CS1	Greece	●	●	●	●	●	●
			CS2	Turkey	●				●	●
			CS3	Slovakia	●	●				●
			CS4	Slovakia	●	●		●	●	●
			CS5	Slovakia	●	●				●
			CS6	Slovakia		●			●	●
Black tide: Oil in the water	Chemistry	Lower	DU	Portugal	■					■
			CS1	Portugal	●	●				
			CS2	Hungary	●	●		●	●	
			CS3	Hungary	●	●		●		●
			CS4	Germany	●					
			CS5	Greece	●		●	●		
Collision of an egg	Physics	Lower	DU	Hungary	■	■		■		
			CS1	Hungary	●	●		●		
			CS2	Denmark	●	●		●		
			CS3	UK	●	●		●		
			CS4	UK	●	●		●		
			CS5	Germany	●	●		●		
			CS6	Germany	●	●		●		
Electricity	Physics	Lower	DU	Poland	■			■	■	■

			CS1	Slovakia	●					●
			CS2	Ireland	●			●		●
			CS3	Turkey	●	●	●			
			CS4	Poland	●				●	●
			CS5	Poland	●				●	●
Oranges	Physics	lower	DU	UK	■			■		
			CS1	Germany	●			●		
			CS2	Germany	●	●		●		
			CS3	Hungary	●			●		
			CS4	Poland	●		●	●		●
			CS5	Sweden	●	●		●		
			CS6	UK	●	●		●		
			CS7	UK	●	●		●		
			CS8	Hungary	●	●	●	●		
Food and Food Labels	Biology	Lower	DU	UK					■	■
			CS1	Turkey	●			●	●	●
			CS2	Hungary			●		●	
			CS3	Ireland						●
			CS4	Portugal				●		●
			CS5	Hungary	●			●		
Global warming	Physics	Upper	DU	Turkey			■			■
			CS1	Denmark			●	●	●	
			CS2	UK				●		●
			CS3	UK			●			●

			CS4	Belgium			●	●		●
Household vs natural environment	Chemistry	lower/upper	DU	Poland	■	■		■		■
			CS1	Ireland	●	●				
			CS2	Greece	●		●			
			CS3	Portugal		●		●		
			CS4	Poland	●	●	●	●		●
			CS5	Poland	●				●	●
			CS6	Poland	●	●	●	●	●	
Light	Physics	lower	DU	Ireland	■		■	■		■
			CS1	Ireland	●	●	●	●		●
			CS2	Ireland	●		●	●		●
			CS3	Greece	●		●	●		●
			CS4	Slovakia			●	●	●	●
Natural selection	Biology	Upper	DU	Denmark	■		■	■		
			CS1	Poland	●		●	●	●	
			CS2	Hungary	●		●	●		
			CS3	Denmark			●	●		
			CS4	Sweden	●		●	●		
			CS5	Hungary	●		●	●		
Plant nutrition	Biology	Lower	DU	Slovakia	■				■	
			CS1	Slovakia	●			●	●	
			CS2	Slovakia	●			●	●	
			CS3	Portugal	●	●		●	●	
			CS4	Hungary	●		●		●	

			CS5	Hungary	●		●		●	●
			CS6	Sweden	●					
Polymers	Chemistry	Upper	DU	Slovakia	■			■		
			CS1	Ireland	●	●	●	●	●	●
			CS2	Poland	●			●		
			CS3	Slovakia				●		
			CS4	Slovakia			●			●
			CS5	Turkey	●	●		●		
Reaction rates	Chemistry	Lower	DU	Ireland	■			■		
			CS1	Hungary	●		●	●		
			CS2	Ireland	●			●	●	
			CS3	UK	●			●	●	
			CS4	Turkey	●			●	●	
			CS5	Germany	●			●	●	
Speed	Physics	Lower	DU	Germany	■			■		
			CS1	Turkey	●				●	
			CS2	Ireland	●			●		
			CS3	Portugal	●					
			CS4	Germany	●					
The proof of the pudding	Chemistry	Lower	DU	Hungary	■	■	■	■	■	■
			CS1	Ireland	●		●	●		
			CS2	Slovakia	●		●	●	●	
			CS3	Greece	●	●	●			
			CS4	Hungary	●		●	●		

Ultraviolet radiation	Physics	Upper	DU	Sweden	■					
			CS1	Denmark	●	●				
			CS2	UK	●			●		
			CS3	Germany	●	●				
Up there… how is it?	Physics	Upper	DU	Portugal	■	■		■	■	■
			CS1	Portugal	●					
			CS2	Slovakia	●			●		
			CS3	Sweden			●	●		
Which is the Best Fuel?	Chemistry	Lower	DU	Turkey	■			■		
			CS1	Turkey		●	●	●		
			CS2	Poland	●	●				
			CS3	Greece	●	●		●		
			CS4	Denmark	●	●	●	●		
Wood lice	Biology	Lower	DU	Sweden	■	■				
			CS1	Sweden	●	●				
			CS2	Poland	●	●				
			CS3	Ireland	●	●				
			CS4	Slovakia	●		●			
			CS5	Portugal		●		●		

Note: Black squares indicate the unit developers' intention to encompass the assessment of different inquiry skills in the draft Unit (DU).
Black dots indicate which inquiry skills were encompassed in the actual case study (CS).

4. Illustrative examples of assessment strategies from classroom practices

In this section, the knowledge domains and skills are further elaborated and illustrated by examples distilled from real classroom case studies. Altogether 19 SAILS units are involved in this report. All these units are based on an original lesson unit, and this unit was adapted and trialled in at least three countries (from which at least two are different from the country that synthesized the experiences from the case studies). By June, 2015, which is the closing month of the school year in most participating countries in the final year of project, 19 SAILS units have been synthesised and are presented in D3.3.

Each SAILS unit is linked to a set of inquiry skills that are especially in the focus of that unit. Table 2 presents the list of the assessed skills in the SAILS units which informed this report. In Table 2, besides the four highlighted inquiry skills, also scientific literacy and scientific reasoning are involved, although they represent different dimensions of the scientific cognitive learning outcomes.

Table 2 shows that repeatedly the case studies contained descriptions on the assessment of inquiry skills different from the skills that were originally intended to be encompassed by the unit developers. This apparent inconsistency can easily be resolved. First, essentially all SAILS units are suitable in improving and assessing any inquiry skills. The unit developers' intention to designate one or more inquiry skills helped the project consortium to select as much representative set of SAILS units as possible for illustrating the diversity in assessment practices and tools. Second, due to cross-cultural and cross-country differences, some case studies did prove to be an illustrative sample for the assessment of a skill different from the unit developers' original intention.

In order to maximize the representativeness of the set of SAILS units, unit developers and the coordinators of the case studies were asked to indicate: the assessment of which inquiry skill can be best illustrated by a given unit or case study. These noteworthy opinions were taken into account when the examples of this current report were selected and analysed. In the following section, the SAILS units (including the case studies belonging to them) serve as the basis for selecting illustrative examples.

Since the current report focuses on the assessment of inquiry skills, the selection process can be understood as if reading Table 2 vertically, i.e., the four inquiry skills, scientific literacy and scientific reasoning constitute the structure of our analysis, and illustrative examples from different rows (from different SAILS units) are selected.

4.1 Curricular content

Within each SAILS unit, curriculum content is taught using an inquiry approach. Curricular content is addressed at different levels. In all case studies, a basic scientific concept is taught in a way that includes some inquiry. The following examples provide some hints about the content-bound development of scientific knowledge.

The *Speed* unit primarily focused on developing the scientific concept of speed. Speed as the ratio of distance and time might remain a formula in many students' mind, but conceptual development may be promoted through investigations designed by students. Case studies conducted by the consortium partners show that students, through inquiry, were able to build a coherent mental representation of three intertwined concepts: time, speed and distance.

Content elements that are or may be involved in curricula, albeit sometimes not of central importance are often evoked by means of brainstorming or mind mapping techniques. In the

Electricity unit, prior to the activities, a teacher chooses a group of learners he or she wants to assess during all brainstorming moments during this particular lesson. It is suggested not to exceed the number of 6 students. During each brainstorming session a teacher checks an appropriate box in the table below (Table 4) to record the frequency of selected students' responses.

TABLE 4 CHECKBOX USED IN ELECTRICITY UNIT

Student	**Brainstorm No 1***		
	prior knowledge	engagement	creativity
Name 1			
Name 2			
Name 3			
Name 4			

The *Food labels* unit provided ample opportunities (and this was utilized by means of a brainstorming technique) to recapitulate the conceptual network necessary to talk about food labels. The scientific concepts of energy and the biological (and everyday) semantic network of nutrient, food, junk food were discussed.

In the *Ultraviolet radiation* unit, the factual knowledge on whether water protects from UV radiation served as a preliminary activation of everyday knowledge.

Other content elements may play a marginal role in the curricula; however as "raw material" for experiments they may evoke students' real-life experiences. Nevertheless, these content elements assist the process of development of inquiry skills built around them. For instance, the *Woodlice* case studies focused on biological facts not emphasized elsewhere in the curricula.

Teachers' lesson plans usually contain scientific concepts to be taught during that lesson. The instructional approach applied during inquiry-based science lessons supports conceptual understanding, and provides opportunities to content-bound ability development.

4.2 Inquiry skills

4.2.1 Developing hypothesis

What does a "good hypothesis" look like? Characteristics can be dependent on the actual content of the task, and one formal feature may sometimes be merely the number of hypotheses developed. Basically, when assessing this inquiry skill, three questions should be answered. Each of the three sub questions can be assessed separately and together. This assessment can take place during the activity, e.g. as students are discussing the questions or examining what they have written in-class or afterwards. Things to look out for:

1. Is the question related to the hypothesis, clear, qualified, is the question testable and specific enough?
2. Is the hypothesis linked to the question? Does it suggest an outcome to the investigation?
3. Is the hypothesis justified, for example based on prior knowledge, existing evidence, students' own observations, or trials?

Students' hypotheses may be generated in response to a question posed by the teacher, as the starting point of an investigation or from students' questions. Table 5 presents an overview of what assessment methods were used by the teachers to assess students' skills of developing hypothesis.

All teachers encouraged student-student and student-teacher dialogue and relied on classroom observation and making judgement based on classroom dialogue as an assessment strategy. Students were facilitated to devise their own material and complete worksheets as assessment artefacts in all disciplines, as well as using other forms of media in the Biology and Chemistry Units. Peer and Self-Assessment was also regularly used in all disciplines.

TABLE 5 ASSESSMENT METHODS USED FOR DEVELOPING HYPOTHESES

Teaching and Assessment Unit	Assessment Method							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self-Assessment	Worksheets	Students devised materials	Presentations	Other
Food and Food Labels	●	●	●		●		●	
Natural selection	●	●						
Plant nutrition	●	●		●	●	●	●	
Wood lice	●	●		●	●	●	●	
Acids, bases, salts	●			●	●			
Black tide: Oil in the water	●	●	●			●		
Household v natural environment	●	●			●	●		
Polymers	●	●		●	●			
Proof of the pudding	●	●			●			
Reaction rates	●	●	●		●	●		
Which is the best fuel	●	●	●		●	●	●	
Collision of an egg	●	●				●		
Electricity	●	●	●	●	●		●	
Floating Oranges	●	●	●	●	●		●	
Global Warming		●						
Light	●	●			●			
Speed	●	●			●	●		
Up there… how is it?	●	●			●			
UV radiation	●	●	●				●	

In the first case study (Ireland) of the *Woodlice* unit, four-level rubrics were used to assess the developing hypothesis skill (Table 6). These four levels have been widely adapted and trialled in the SAILS project. In this actual unit three variables (food, light and moisture) seem to play their important role in woodlice life. Naturally, the experimental design can either precede or follow the hypothesis formulation.

TABLE 6 RUBRIC USED IN WOODLICE CS3

category	Emerging	developing	Consolidating	extending
Formulating hypotheses	A prediction is Made	A testable prediction is made linked to the question	A testable prediction to the question is made that suggests a clear outcome	A testable prediction to the question is made that suggests a clear outcome based on scientific reasoning

The second case study (Portugal) of the Plant nutrition unit applied the same four-level instrument for assessing the formulating hypothesis skill, but with alternative descriptions (Table 7). It is interesting to compare how the same labels are defined in this case.

TABLE 7 RUBRIC USED IN PLANT NUTRITION CS3

category	Emerging	developing	consolidating	Extending
Formulating hypotheses	Formulates hypotheses that are not consistent with the planning or that are not eligible for investigation.	Formulates hypotheses that are consistent with the planning of the experiment.	Formulates hypotheses that are consistent with the planned experiment and are based on the research questions.	Formulates hypotheses that are consistent with the planed experiment. Those hypotheses are based on the research questions and identified variables.

Obviously, both types of descriptions can be used. It is the teachers who will find them usable and easily applicable in classroom situations.

The first case study (Hungary) of the Collision of an egg unit provided great opportunity for students to develop their own hypotheses. The level of their hypotheses has been assessed according to three-level rubrics, where the levels are defined according to how students' statements are scientifically justified (Table 8). The rubrics used were accompanied by helpful teacher questions which may support students in improving the level of their answers.

TABLE 8 RUBRIC USED IN COLISSION OF AN EGG CS1

Helpful questions:			
What do you expect to happen? Why does it happen? Can you explain how your hypothesis follows from what you have learnt?	The student formulates the hypothesis but is unable to explain it.	The student formulates the hypothesis and is able to explain the hypothesis with the help of questions.	The student explains the hypothesis and supports it with scientific facts.

In the original *The proof of the pudding* unit, students had to make some initial predictions about the desired ratio of the pudding components. The quality level of this inquiry skill was assessed by means of three-level rubrics. For this, both the observed classroom discussion and students' answers to a questionnaire were used. On the other hand, in the second case study (Ireland), five-level rubrics were used. The scale they used (1 for the highest, and 5 for the lowest level performance) indicate historical cross-cultural differences in the marking systems of different countries. Whether on a five-point scale the bigger numbers are associated with better or worse performance can be traced back to the differences in using school marks.

4.2.2 Planning and implementing investigations

Planning investigations as an inquiry skill is addressed by almost each SAILS unit. According to Walker (2007), different "inquiry levels" can be identified. When the phase of formulating research question and planning investigations is initiated by students, this can be described as the highest level of inquiry, and this proves to be the most difficult to accomplish. The planning investigations inquiry skill is a complex skill, the assessment of which can be accomplished in several ways.

The table (Table 9) presents an overview of what assessment methods were used by the teachers to assess students' skills of planning investigations. 18 units involved classroom dialogue as an assessment method and 17 units employed teacher observation as part of the assessment strategy. Students were facilitated to devise their own materials and complete worksheets as assessment artefacts in all disciplines. Peer and Self-Assessment was also employed in all disciplines.

TABLE 9 ASSESSMENT METHODS USED FOR PLANNING INVESTIGATIONS

Teaching and Assessment Unit	Assessment Method							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self-Assessment	Worksheets	Students devised materials	Presentations	other
Food and Food Labels	●	●	●	●	●	●		●
Natural selection	●	●			●			
Plant nutrition	●	●		●	●	●	●	
Wood lice	●	●				●	●	
Acids, bases, salts	●							
Black tide: Oil in the water	●	●	●		●	●		
Household v natural environment	●	●	●		●	●	●	
Polymers	●	●		●	●			
Proof of the pudding	●	●			●			
Reaction rates	●	●	●		●	●		
Which is the best fuel	●	●	●	●	●	●	●	
Collision of an egg	●	●	●	●		●		
Electricity	●	●	●	●	●	●		●
Floating Oranges	●	●	●	●	●	●		●
Global Warming		●						
Light	●	●			●			
Speed	●	●			●	●		
Up there… how is it?	●	●			●	●		
UV radiation	●	●	●			●	●	

The first case study (Portugal) of the Black tide unit applied an integrated assessment instrument for three subskills (Table 10).

TABLE 10 RUBRIC USED IN BLACK TIDE, OIL IN WATER CS1

Subskills	Performance levels		
	1	2	3
Define strategies and procedures	Does not define the necessary strategies and procedures to achieve its goal.	Defines with some difficulty the necessary strategies and procedures to achieve its goals.	Defines the necessary strategies and procedures to achieve its goals.
	Unclear planning requiring reformulation.	Planning well-presented but requiring reformulation.	Clear, concise and complete planning.
Know resources and chose them adequately	Does not select adequate resources according to the goals and strategies.	Selects some resources that are adequate to the goals and strategies.	Selects the resources that are adequate to the goals and strategies.

Sub-skill: decide what you want to do to find out the answer to the question, including identification of variables and consideration of fair testing

The core element of this skill component is the identification, definition and separation of different variables in the experimentation process. One crucial element is whether the student is capable of separating the independent and dependent variables (from 7^{th} or 8^{th} grade, the exact labelling of these variables might be feasible). The importance of distinguishing all relevant variables in the experiment can be the basis for assessing the skill of planning investigations. In a previous phase of the project, four level rubrics were used as a generalizable idea how this skill may be assessed either during classroom discussion or in post-hoc questionnaires.

TABLE 11 RUBRIC USED FOR ASSESSING STUDENTS SKILLS IN IDENTIFYING VARIABLES IN AN INVESTIGATION RELATING TO ELECTROLYSIS

0	No answer
1	Mention concepts from the actual experiment (temperature, plate, etc.)
2	Explicitly state all variables
3	Explicitly state all variables changed and measured

This four-level scale can generally be used to assess whether a student is capable of identifying variables in a research problem (Table 11). The first level is a big step towards planning, since relevant scientific concepts are mentioned. The second step indicates that these scientific concepts can be handled as variable, while at the highest level variable characteristics are understood (kept constant, changing, to be measured, etc.)

Several other examples of assessing this inquiry skill were provided and described in other case studies. In the *Proof of the pudding* unit, students were required to produce an edible mixture from its two basic components. Students had to decide how to change the ratio of the components. Another variable was the time of cooking. Students' skills of keeping one of these variables constant while manipulating the others have been assessed by means of a three-level rubric. The basic schemata for three-level scoring are: wrong, partially right, right. In IBSE assessment contexts, it is often easy to decide whether an inquiry skill failed to work, partially worked or smoothly worked in a given task. Therefore three-level rubrics enables for making quick decisions supporting both formative and summative assessment purposes.

Some case studies gave students a more open problem where there were many potential variables and it proved to be even more difficult to assess the quality of this skill. In the *Collision of an egg* unit, students were free to choose dependent variables. One group kept the height of falling constant, and varied the surface while another group decided to try out three different heights with different surfaces. In this latter case not all combinatorial possibilities were consistently tested. Here the three-level assessment scale did not follow the previous wrong-partially right-right scale, but was an ordinal scale measure of the planning skills, similarly to the aforementioned 0-1-2-3 scale.

The teacher provided oral feedback (without using rubrics) also in the fourth case study (Germany) of the *Black tide* unit. During that lesson formative assessment has been implemented. The teacher used it in a traditional and very economical way: the very small number of students allowed her to watch all groups carefully, to listen to the discussion in the groups and to give specific advice to the group with regard to the experiments. Rubrics were not used and the formative assessment was given on the spot. Peers shared their ideas and commented on the ideas of the other groups.

There is a concise opinion from a teacher in Germany (third case study of the *Ultraviolet radiation* unit) about the applicability of rubrics:

> "The rubric system of the UV radiation unit was shown to the teacher ahead of the lesson. She told that she was unable to use the rubrics because she had no time to allocate students in the rubrics during the experimental process. Her conclusion was that the rubrics could on the one hand be used in a team teaching situation (two teachers) or on the other hand must be adapted as a self-assessment tool."

The *Plant nutrition* unit applied the "usual" (widely used across skills and countries in the SAILS project) four-level rubrics to assess different facets of the planning investigation inquiry skill (Table 12).

TABLE 12 RUBRIC USED IN PLANT NUTRITION UNIT

category	**emerging**	**developing**	**consolidating**	**extending**
Layout of samples	Procedure precise, but small distances between samples (10cm)	Indicates chosen method and argues its speed	Indicates chosen method and argues its accuracy	Indicates and compare speed and accuracy of chosen method
Data entry	Data entered into a continuous text of process	The layout is less accurate, time is marked, use a table	Able to reason the procedure in practical terms (for example to use the full length of the table)	Able to reason the procedure, builds on the fundamental of photosynthesis

Similarly, in the *Ultraviolet radiation CS1*, students were asked to investigate UV radiation, and they, rather than the teacher, decided whether a lamp in a room or the sunlight would be taken as the source of radiation. The teacher was previously provided with rubrics as assessment tool, but restricted herself to judging students' planning skills based on discussions with students.

The *Woodlice CS1* allowed students to investigate the living conditions of woodlice provided evidence of rather different levels of planning skills. Three different levels were identified, each illustrated in the case study report. One interesting feature of a "mid-level" plan was to add new variables to the experiment inconsistent with the research question instead of eliminating or fixing some of them.

Distinguishing dependent and independent variables is of crucial importance in understanding the planning phase of scientific inquiry. The *Fish-eating bird CS1* provided data about students' difficulties. To the teacher's question "*What were the experimental and the control conditions?"* students often wrongly thought the two phases of the bird's movement in the first experiment were relevant. In some instances the student described the experimental and the control conditions, but did not indicate which was which.

The *Oil in the water* and the *Up there how is it* (CS1) used a three-level rubric to assess the skill of defining the goals of the investigations, with the degree of clarity used in students' descriptions being the defining characteristic used in judging performance. The *Martian bacteria DU*, however, use a four-level rubric assessment scale. The lowest level of the four-level scale can be understood as the absence of the inquiry skill, while the highest level refers to a holistic approach in identifying and operationally defining all relevant variables.

The flexibility in planning was highlighted in the *Floating orange CS1*, where students were encouraged to modify their inquiry questions while going through the inquiry.

Considering two parts of planning and implementing investigations:

(a) Decide what materials you need

In the *Collision of an egg unit,* students were free to choose their equipment and materials is the where an extensive range of equipment was available. The teacher even allowed the possibility for students to request additional material. Four level rubric shown for Electricity unit (Table 13).

TABLE 13 RUBRIC FROM ELECTRICITY UNIT

Assessed Skill	Emerging	Developing	Consolidating	Extending
Planning investigation of conducting properties of different materials	The student... ... lists a limited number of objects made of 1-2 different kinds of materials but does not write the plan at all or the investigation plan is incomplete	The student... ... lists a limited number of objects made of 1-4 different kinds of materials and the investigation plan is almost correct	The student... ... lists a limited number of objects made of over 4 different kinds of materials and the investigation plan is almost correct	The student... ... lists a limited number of objects made of over 4 different kinds of materials and the investigation plan is complete

(b) Decide how to record the information

In many draft units teachers' assessed students' skills of recording information. The quality of data collection was scored by a trichotomous item as shown in Table 14 below.

TABLE 14 RUBRIC USED IN WOODLICE CS2

2 points level	4 points level	6 points level
Student can interpret data correctly (categorizing the measured variables as lesser – greater) but cannot create a proper graph based on them	Student can present the data on a graph, but the graph lacks or has poorly developed elements as axes titles, scale, legend etc.	Student can present the data on appropriate graph(s) having all necessary elements as axes titles, scale, legend etc. prepared correctly
Student can point out basic / selected sources of biased / incorrect results of the experiment	Student can enumerate all main factors that might be sources of biased/incorrect results of the experiment	Student can analyse all main factors that might be sources of biased/incorrect results of the experiment and indicate ways to avoid them in the future
Student can propose elements of a method serving to improve the experiment	Student can propose improvement of the course of the entire experiment step by step	Student can compare results of other groups, discuss data interpretation and propose methods to improve both own and the other groups' experiments

4.2.3 Working collaboratively

Both the quality of the discussion in a group work context and the degree of collaboration was assessed in the SAILS CSs. Assessment of working collaboratively was conducted primarily through teacher observation and classroom dialogue after which immediate oral feedback was given to students. Students were facilitated to complete worksheets in all disciplines but devise their own material and as assessment artefacts only in 4 units (3 Chemistry, 1 Biology). Furthermore, students' self- and peer-assessment proved to be usable for the assessment of this inquiry skill. The table (Table 15) presents an overview of what assessment methods were used by the teachers to assess students' skills of working collaboratively.

The use of peer- and self- assessment strategy is illustrated by the *Electricity* unit, in the fourth case study (Poland), where the teacher creatively implemented two series of questions for students first about themselves (Table 16), then about their peers (Table 17)Both assessment tools were used at the end of the unit.

TABLE 15 ASSESSMENT METHODS USED FOR WORKING COLLABORATIVELY

Teaching and Assessment Unit	Assessment Method							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self Assessment	Worksheets	Students devised materials	Presentations	other
Food and Food Labels	●	●	●		●		●	
Natural selection	●	●						
Plant nutrition	●	●		●	●	●	●	
Wood lice	●	●		●	●	●	●	
Acids, bases, salts	●			●	●			
Black tide: Oil in the water	●	●	●			●		
Household v natural environment	●	●			●	●		
Polymers	●	●		●	●			
Proof of the pudding	●	●			●			
Reaction rates	●	●	●		●	●		
Which is the best fuel	●	●	●		●	●	●	
Collision of an egg	●	●				●		
Electricity	●	●	●	●	●		●	
Floating Oranges	●	●	●	●	●		●	
Global Warming		●						
Light	●	●			●			
Speed	●	●			●	●		
Up there… how is it?	●	●			●			
UV radiation	●	●	●				●	

TABLE 16 RUBRIC USED IN ELECTRICITY CS4

Task for learners: Reflect on your involvement in a team work during the lesson. Using the scale 0 (not at all) to 6 (very much), score your engagement, according to the statements listed in the table below.			
Item	*Peer 1*	*Peer 2*	*Peer 3*
I was involved in planning the experiment.			
I carried out the given tasks.			
I helped colleagues.			
I was involved in collection of data.			
I was active in performing the experiment.			
I communicated properly with the others.			

TABLE 17 RUBRIC USED IN ELECTRICITY CS4

Reflect on involvement of the peers from your group in a team work. Using the scale 0 (not at all) to 6 (very much), score their engagement, answering the questions listed in the table below.			
Question	*Peer 1*	*Peer 2*	*Peer 3*
Did your colleague take part in planning the experiment?			
Did your colleague take part in carrying out the given tasks?			
Was your colleague helping the group?			
Was your colleague engaged in data collection?			
Did your colleague take part in performing the experiment?			
Did your colleague communicated properly in the group?			

Since in Poland a six-level grading scale is used in the schools, the numbers used were analogous with this system. Other teachers may prefer a non-numerical scale that emphasizes the positives of the work done rather than a potentially ego-focussed numerical scale,

Another nice example of how self-assessing the quality of group work can be realized was trialled in the second case study (Poland) of the *Polymers* unit. The frequency of different activities was judged by the students on a four-point scale. Some items of the questionnaire are presented below as illustration (Table 18).

TABLE 18 RUBRIC USED IN POLYMERS CS2

Workgroup assessment – Self-assessment sheet

Specify how often *each* situation occurred when working in a group*:*
In your assessment use the scale below:

1	*2*	*3*	*4*
Hardly ever	*rarely*	*sometimes*	*often*

Workgroup assessment		1	2	3	4
1	I took part in the discussion				
2	I listened carefully to what other students were saying				
3	I offered suggestions for the solution and other members of the group accepted them				
4	I agreed to the suggestions of my friends				

The Food and food label unit provided opportunity for group discussion on the question:

- What do you think junk food is?

Each group completed a placemat with 4 sections. Each student had a segment to write in on the placemat and then collectively the group decided on a consensus definition for the top segment – this promoted small group debate. Teacher formed the groups - based on his/her knowledge of the student (quiet children together – more dominant students together) in participating in group discussions. Groups were not formed on the basis of academic ability or gender.

The teacher then circulated between the groups asking probing questions to individuals/groups to encourage students to decide on an appropriate scientific definition and encourage *appropriate contributions from everyone*. Feedback was given in whole class discussion.

Similarly, a four-level rating scale was applied in the second case study (Hungary) of the Natural selection unit. In that case study, a combined formative assessment of two inquiry skills took place. The assessment of the working collaboratively and forming coherent arguments skills was obtained according to the following criteria (Table 19).

TABLE 19 RUBRIC USED IN *NATURAL SELECTION CS5*

	Emerging	Developing	Consolidating	Extending
Working collaboratively and forming coherent arguments	Participates in group work but with interruptions. Carries out the tasks given to them, but does not volunteer to do things. Generally participates in group debates as a passive observer.	Participates in group work without interruptions but with varying intensity. Occasionally volunteer to do tasks. Expresses opinions in debates but does not present coherent or persuasive arguments.	Participates in group work actively and without interruption. Often volunteers to do jobs as part of the team. Actively participates in debates and supports opinions with arguments.	Tends to have a leading role in group work, efficiently organizes and assists work of peers. Brings up persuasive arguments in debates, is able to appreciate others' points of views and can be convinced to change mind if presented persuasive arguments.

Who else than the students themselves can genuinely judge the appropriateness of their teamwork? In the first case study (Germany) of the *Floating orange* unit students were provided with self-assessment questionnaires (Table 20).

TABLE 20 RUBRIC USED IN *ORANGES CS1*

Item	I achieve this goal totally	I achieve this goal partly	I don't achieve this goal
I did let my schoolmates finish their argumentations and did not disrupt them.			
I did not do inappropriate comments to my schoolmates' argumentations.			
I did not put my schoolmates under pressure or force them to do what I wanted.			
I did inform all group members about planed investigations or upcoming inquiry processes.			

Of course, the data coming from self-assessment questionnaires is worth being validated against teacher (or other) observations.

It is possible for all of the methods described above to also consider gender issues. To do this teacher and students can both be asked about the boy-girl interactions and discuss if their perceptions were the same as each other. A variety of means can be used on which to base these discussions including verbally, by writing notes, or using pro-formas. Through discussions of comparisons of feelings students can learn to empathise.

4.2.4 Forming coherent arguments

Two frequent modes of assessing this skill were evident in the case studies. Written assignments revealed some strengths and weaknesses of students' argumentation. The other form used involved teacher/student discussions where the teacher provided immediate oral feedback where there was incoherence in the argumentation (Table 21).

TABLE 21 ***ASSESSMENT METHODS USED FOR FORMING COHERENT ARGUMENTS***

Teaching and Assessment Unit	Assessment Method							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self-Assessment	Worksheets	Students devised materials	Presentations	other
Food and Food Labels	●	●			●			
Natural selection	●	●			●	●		
Plant nutrition	●				●		●	
Wood lice	●					●		
Acids, bases, salts	●		●		●			●
Black tide: Oil in the water	●	●	●	●				
Household v natural environment	●	●			●		●	●
Polymers	●	●		●	●			
Proof of the pudding	●	●	●		●			
Reaction rates	●	●	●		●	●		
Which is the best fuel	●	●	●		●		●	
Collision of an egg	●	●				●		
Electricity	●	●			●			
Floating Oranges	●	●			●			
Global Warming	●	●	●	●	●	●	●	
Light	●	●			●			
Speed	●	●			●	●		
Up there… how is it?	●	●			●	●		
UV radiation	●	●	●				●	

In the *Proof of the pudding* unit, both types of these assessment strategies have been observed. In the first case study (Greece), students' written worksheets served as the basis for even summative assessment. In the second case study (Ireland), however, not even rubrics were used, but teachers watched how the members of groups collaborated. During the activity oral formative feedback was provided by the teacher. Furthermore, during the peer assessment, students listened to their classmates' arguments. Interestingly, even within the same case study (CS3, Slovakia) different assessment strategies were used for two different inquiry skills. For assessing the formulating hypotheses skill, the teacher filled in rubrics, whereas for assessing the forming coherent arguments skill, students conducted peer-assessment, following the teacher's explanation on the criteria of the rubrics.

In the *Light* unit, three-level assessment criteria were suggested in the original unit description, and in different case studies this suggestion was followed. The teacher in the third case study (Greece) used the three level assessment criteria described in the unit for "Interpreting results and drawing conclusions" to make judgements on the student's abilities of forming coherent arguments. However, the teacher in the first case study (Ireland) described and used different three-level criteria for making judgements on the skill of forming coherent arguments, and did this both for written responses on worksheet as well as making judgements on verbal responses. The three levels are described as follows:

1. The student does not provide and / or does not explain the arguments in his/her own words (construction); key arguments aren't properly developed.
2. The student presents and explains her arguments, explaining the key arguments but not completely. In case of verbal communication, this level includes complete answers obtained only after prompting by the teacher.
3. The student presents and explains his/her arguments in his/her own words (construction), properly developing the key arguments.

The forming coherent arguments skill is very suitable for peer-assessment strategies. In the first case study (Greece) of the Acids unit, a four-level ordinal scale was used (Table 22). Students had to judge the quality of their peers' answers.

TABLE 22 ***RUBRIC USED IN ACIDS AND BASES CS1***

	Excellent (4)	Good(3)	Needs Improvements (2)	Needs Reconsideration (1)	Score
Do they use arguments in order to convince you;					
Is the argumentation being used complete;					
Does the argumentation being used feel right?					
Does the answer seem right?					
				Total:	

The *Global warming* unit provided ample space for arguing for and against different statements. Both in the UK and Danish case studies, students used rubrics to assess imagery students' (Students A and B) opinions. As observed in the Danish case study, students did find difficult to use the language prefabricated for them in the rubrics.

Identify the parts of the graphs that do not support the conclusion by Student A and present supportive arguments for the conclusion made by Student B. Use the rubric to check your answer (Table 23).

TABLE 23 RUBRIC IN *GLOBAL WARMING* UNIT

	Level 1	Level 2	Level 3
Skill assessed: Using scientific information	Makes reference to both graphs (as a whole).	Makes reference to both graphs (as a whole and in detail).	
	Presents supportive arguments for at least of one of the student's conclusions.	Presents supportive argument/s for both of the student's conclusions.	Presents several supportive arguments for both of the student's conclusions.
	Attempts to provide scientifically reasonable justifications for arguments.	Provide scientifically reasonable justifications for arguments.	Provide scientifically valid justifications for arguments.

As seen from this example, inquiry-based science education has the inherent potential to improve not only inquiry-related knowledge components, but the ability to reflect on other people's thoughts. These metacognitive knowledge components seem to be especially valuable and transferable.

4.3 Reasoning skills and abilities

The SAILS units are suitable for diagnosing the efficient and appropriate functioning of different inferential rules in students' minds. All implementations focussed on assessing reasoning skills and abilities and had a wide range of characteristics as shown in Table 24.

Teachers adopted a range of approaches to gather evidence of student learning in order to make a judgement of their scientific reasoning abilities (Table 24).

The Electricity unit requires students to draw mind maps containing their scientific and everyday concepts in the field of electricity. The quality of such mind maps can be assessed by means of four-level rubrics (Table 25):

TABLE 24 RUBRIC USED IN *ELECTRICITY* UNIT

task	emerging	developing	consolidating	extending
Drawing a mind map	Student's mind map is empty or full of inadequate words, for which the student cannot describe a relation to electricity	Student is able to... ... draw a mind map containing only a few words and/or the words are listed with no relation to each other	Student is able to... ...draw a mind map with more than 10 words, both scientific and belonging to a common language, but the visualization of the relationships and categories is poor	Student is able to... ...draw a mind map with more than 10 words, both scientific and belonging to a common language, with a good visualization of the relationships and categories

TABLE 25 ASSESSMENT METHODS USED FOR REASONING SKILLS

Teaching and Assessment Unit	Assessment Characteristic							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self Assessment	Worksheets	Students devised materials	Presentations	other
Natural selection	●	●				●		
Plant nutrition	●	●		●		●	●	
Wood lice	●	●				●		●
Acids, bases, salts	●				●			●
Black tide: Oil in the water	●	●			●			
Household v natural environment	●	●			●		●	●
Polymers	●	●		●	●			
Proof of the pudding	●	●			●			
Reaction rates	●	●	●		●	●		
Which is the best fuel	●	●	●		●	●	●	
Collision of an egg	●							
Electricity	●	●			●	●		
Floating Oranges	●							
Global Warming	●	●	●	●	●	●	●	
Light	●	●			●			
Speed	●	●			●	●		
Up there… how is it?	●	●			●	●		
UV radiation	●	●	●				●	

The four descriptors described in Table 25 can easily be adapted for different learning settings or classrooms. The developing understanding is most easily observable, and more advanced understanding is described in terms of the level of visualization. In a case study in Poland, the teacher extended the four descriptions to six columns making it similar to the six level grading system used in their school systems and helped them to integrate their inquiry approaches into existing systems.

The teachers that used the Global warming unit found that it provided the opportunity to assess how students were able to give both arguments and counter-arguments. The assessment approach used in the case study from the United Kingdom took the form of peer-assessment where students used post-it notes to give feedback.

A further assessment strategy used to assess scientific reasoning was documented in the Household unit describing how students, working in their inquiry groups made presentations about their process and outcomes including their conclusion and critical reflections. This not only gave the opportunity for students to articulate their understanding for their peers to review but it also made their learning explicit for the teacher to draw on when assessing their current understanding of this specific inquiry skill. It was also used formatively and informed the teacher's planning of future lessons to develop their skills in a specific direction.

The Acids unit gave an opportunity to assess students' reasoning ability and how to make inferences. The students were required to consider and justify their views on whether or not a given solution was an acid or base and explain how this can be deducted from the colour of the pH paper. The way the students grouped materials according to whether they are acid, bases or salts gave an opportunity for the teacher to assess other reasoning abilities. These inquiry skills were both summatively assessed and formatively assessed through a range of approaches including the way the teacher gathered evidence about how a group of students fulfilled the previously agreed criteria about the quality of categorization and making well founded inferences. In a number of instances the assessment criteria were shared with the students at the start of the inquiry and on occasions the students also helped to craft the wording in the descriptors. This helped the learning and assessment process.

The Plant nutrition unit also used a four column matrix / rubrics like the Electricity unit (Table 26). The progression is also described from the left towards the right and gives descriptive guidance on what achievement would look like as an accomplishment. While it is broadly similar in approach to the Electricity unit and describes progression from emerging towards extending, the actual descriptors have been refined to reflect achievement specific to photosynthesis. This then made it possible for it to be used by pupils to self and peer review, as well as the teacher, who could also tease out further assessment evidence where a judgment was not yet possible as further information was necessary.

TABLE 26 RUBRIC FROM *PLANT NUTRITION* UNIT

	Emerging	Developing	Consolidating	Extending
Thinking about photosynthesis based on enrolment and formulation of conclusions	Understanding the procedure (Example: When we do it this way, we see the colour change of indicator)	Arguments show understanding of the procedure (Example: The colour change of indicator occurs as the result of different distances from light)	Arguments show understanding of the process (Example: The colour change of indicator occurs as the result of photosynthesis)	Arguments points to the understanding of the purpose of experiment and the principle of action. (Example: We achieved higher concentration of carbon dioxide because lack of photosynthesis by decreasing light intensity)

Some experienced IBSE teachers developed the idea of a matrix / rubric' into an 'arrow ' as shown in Figure 3 below. It captures how the student is able to take greater responsibility for their learning and contribute to the assessment process. The skill focused on was 'Reasoning and Argumentation'.

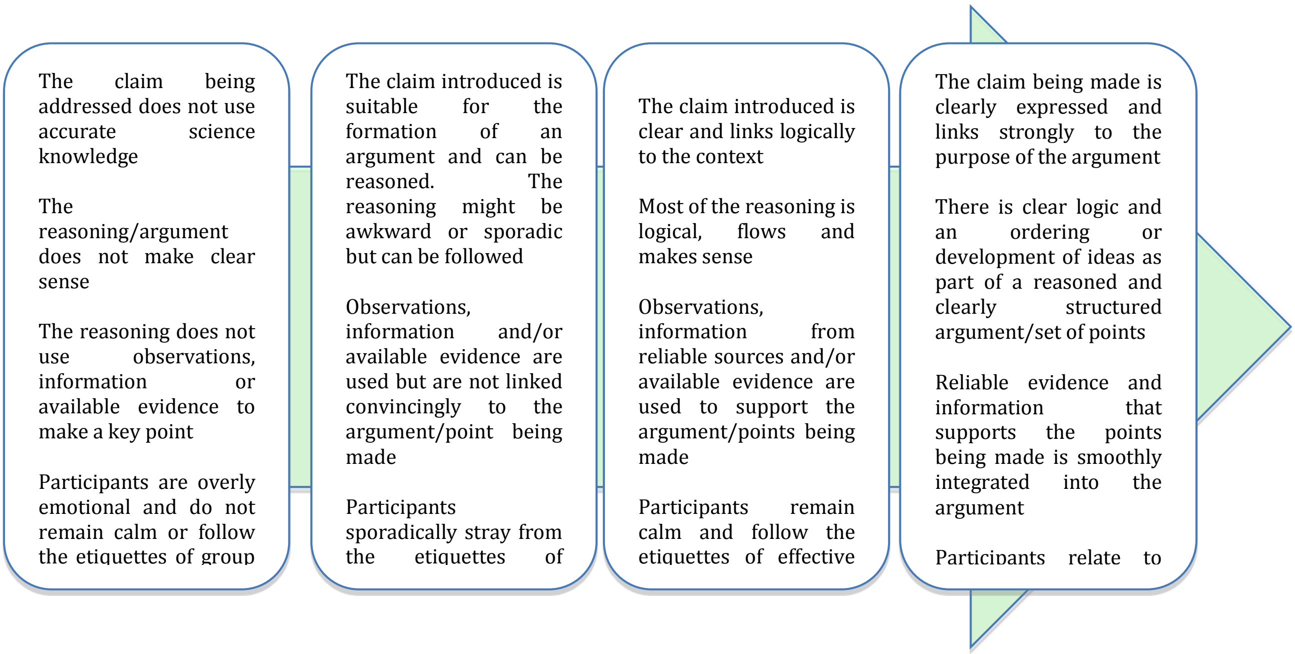

Figure 3 Arrow rubric used to assess reasoning and argumentation

The teacher suggested that this arrow rubric could be used **as** follows:

During a discussion or during argumentation the above ideas might to apply when making reasoned points that either support or challenge the idea or explanation being offered. The aim is that a valid prediction or conclusion is generated and accurately articulates the relationship between the prediction or conclusion and the principle/premise that was used. A valid generalisation might then be made and the logic and reasoning of this generalisation could then be tested through inquiry.

4.4 Scientific literacy

The application of scientific knowledge in different situations or contexts is a widely acknowledged target of science education. Scientific knowledge can be recalled and applied in different situations. Table 27 gives an overview of the wide variety of methods that were employed by the teachers that developed the CSs to assess scientific literacy. Assessment of classroom dialogue was employed in all 4 Biology Units and in 6 of the Chemistry/Physics Units.

TABLE 27 ASSESSMENT METHODS USED FOR SCIENTIFIC LITERACY

Teaching and Assessment Unit	Assessment Method							
	Classroom Dialogue	Teacher Observation	Peer Assessment	Self Assessment	Worksheets	Students devised materials	Presentations	other
Food and Food Labels	●		●					
Natural selection	●	●			●			●
Plant nutrition	●							
Wood lice	●	●				●		●
Acids, bases, salts	●			●				
Black tide: Oil in the water								
Household v natural environment	●	●			●		●	●
Polymers	●	●		●	●	●		
Proof of the pudding	●	●			●			
Reaction rates	●	●			●	●		●
Which is the best fuel	●	●	●		●		●	
Collision of an egg							●	
Electricity	●	●	●	●	●	●		●
Floating Oranges	●		●	●				
Global Warming		●	●	●	●	●	●	
Light	●	●		●	●			
Speed	●	●			●	●		
Up there… how is it?	●							
UV radiation	●	●	●				●	

If science education provides school-context-bound knowledge then students will have less chance of applying their knowledge in everyday situations as compared to science education that builds on problems from different contexts. What is more, even students themselves may and should be aware of the connection between school lessons and everyday problem situations. This kind of awareness can be verbally reported by themselves (metacognition), as in the second case study (Turkey) of the Acids unit. At the end of the lesson unit students filled in a questionnaire with open-ended questions:

- What did I have troubles with during the lesson?
- What did I learn at the lesson?
- What else would I like to learn?
- What do I remember well?
- Where can I use what I did at the lesson?

Which SAILS units can be considered as describing everyday authentic problems? The phenomenon of authenticity is culturally-bound, but in Europe the *Food labels*, the *Oil in the water*, *Ultraviolet radiation* and *Collision of an egg* units seem to be very close to students' interest and activities. When introducing these topics, students feel that they learn something from and about their life.

The assessment of scientific literacy involves questions on scientific inquiry. In other words, inquiry skills are part of scientific literacy when the task is authentic.

The Black tide unit focuses inherently on a topic that is built on a problem selected from the news regularly bombing us: how to prevent and handle nature disasters. Indeed, the question how to find a good cleaning agent to remove oil from the bird feathers, allowing them to survive this kind of environmental disaster (CS3, Hungary), is of multidisciplinary nature: biology, chemistry, physics and even geography knowledge may play a part in the discussion.

Similarly, the Global warming unit addresses a topic that has several connections to everyday experiences. Here experiences refer not to real-life observation of climate tendencies but to the information flood coming from the media, and very often unjustified claims or incorrect chain of inferences are observed. When the title of an article tells "Glaciers melting due to global warming" and at the end of that article the consequence may be: melting glaciers provide further evidence on global warming, there should be an agreement among science educators that providing and justifying arguments and counter-arguments is an essential knowledge component.

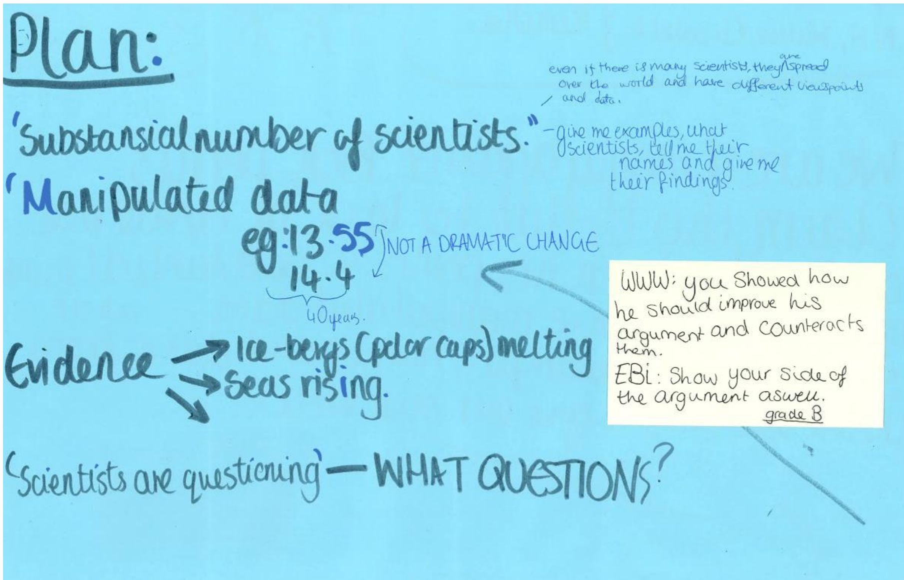

Figure 4 Student poster from *Global Warming* CS2

The assessment of students' posters (Figure 4) allowed peer assessment, and rubrics or other hierarchically structured assessment tools require the teacher's explanation before using them in peer-assessment practices.

5. Conclusions

This report presents the final versions of all of the assessment frameworks and instruments that have been developed as part of the SAILS project. Based on established research into cognition and assessment, it provides illustrative examples of classroom based assessment practices applied across the sciences. The SAILS team identified and selected inquiry activities that promoted these skills and competences and developed assessment strategies appropriate for each skill and competency highlighted in these activities. Within the SAILS project, inquiry in the science classroom is understood to be the intentional process of providing opportunities where students are actively involved in diagnosing problems, critiquing experiments and distinguishing alternatives, planning investigations, researching conjectures, searching for information, constructing models, debating with peers, and forming coherent arguments. In carrying out this project, SAILS has focussed on supporting the development of four inquiry skills (developing hypotheses, working collaboratively, forming coherent arguments, planning investigations) as well as the competencies of scientific reasoning and scientific literacy.

Through a dynamic collaboration between SAILS partners and teachers, nineteen SAILS Inquiry and Assessment Unit*s* have been developed which showcase the benefits of adopting inquiry approaches in classroom practice, exemplify how assessment practices are embedded in inquiry lessons and illustrate the variety of assessment opportunities and processes available to science teachers. The experiences of inquiry and assessment practices in the second level classroom have been collated on over 100 case study reports written by second level teachers that have trialled SAILS units. Teachers have adapted and adopted many different assessment strategies to assess the same skill, as described in the case studies. The case studies provide a narrative of how the teachers approached inquiry in the unit, how feasible the lesson was with the chosen class and how they assessed their students' learning. They also highlight any issues encountered, relating to cultural perspectives and other equity issues, such as gender. It is clear that teachers have adapted SAILS units to also focus on additional skills that the teacher wished to develop. The assessment criteria used were also modified to suit student age and their experience level with inquiry and, in some case studies these criteria were also shared with the students so that they developed their experience with self-assessment and peer-assessment.

Although some case studies indicated the possibility of summative assessment forms (e.g., when scores from different rubrics were summarized and converted to school marks), most of the practices presented serve the aim of formative and diagnostic assessment. Assessing content knowledge often results in a diagnosis about students' semantic networks and misconceptions. Mind mapping and brainstorming techniques enabled students to reveal their prior and prerequisite knowledge on different topics. Assessing reasoning and literacy components resulted in diagnoses on misuse of logical or inductive thinking processes. In these cases, the ongoing assessment helped students to reformulate and improve their answers.

The analysis of the case studies highlight the different ways that inquiry teaching, learning and assessment have been approached in the classroom across the 12 participating countries. Assessment has been shown to be a dynamic and cyclic process that takes place at multiple timeframes. Teachers' assessment practices have been influenced by several factors – e.g. teachers' own experiences, student cohort and local curriculum. The case studies highlight to teachers that a variety of assessment strategies is both necessary and required for assessing inquiry learning in science.

According to teachers' experiences as described in each SAILS unit, usually at most two or three inquiry skills may be assessed within one lesson. This is in line with the SAILS unit typically matched with two or three skills, which are especially well addressed in that unit. Taken into account that in many cases, group work is an important part of the classroom situations, and since assessment may be realized in both individual and group levels, the assessment strategies for inquiry skills require concentrated efforts by the teachers. Nevertheless, the first case study (Ireland) of the Polymers unit provides a nice example where all inquiry skills are integrated, developed and assessed in a collaborative learning environment. In that case study, assessment tools that are completed by both the teacher and the students served as the assessment instruments.

The practice of involving students in the assessment process is strongly promoted, e.g. in the development of criteria for making judgements and through the use of student self- and peer-assessment. There is a tendency in the context of IBSE that formative assessment takes place at the group-level. This ensures that even when using hierarchically structured (ordinal scale) assessment, the scores received are paired with actual performance and not with individual psychological characteristics. The aim of students' self-assessment and peer-assessment practices is to make distinctions between actual performance and long-time traits as aptitude or intelligence.

In conclusion, this report presents the SAILS Framework which describes each of the inquiry skills and competencies considered by the SAILS team and proven strategies for assessing them. Two key characteristics of the SAILS approach have been observed: students are more involved in the active learning process; and students developed lifelong skills critical to thinking creatively, as they learn how to solve and discuss problems using logic and reasoning. SAILS approaches have enabled teachers to both observe what students could do and to hear the reasons why students took certain decisions. It also revealed the range of inferences students made from their data and how students interpreted their results in terms of their scientific understanding. The teachers had more opportunities to assess their students' developing skills and understanding during the inquiry process and reported that it helped them get a clearer view of how students were doing and also what students needed to help them progress. The many examples presented illustrate the value of interplay of focus on individuals, groups, class and serve to inspire teachers to change their inquiry and assessment practices. However, developing inquiry practices, by both teachers and students, takes time as teachers and students need to learn to act, interact and learn in different and often new ways. Teachers are encouraged to facilitate students learning by inquiry, having set up their students to self-assess and monitor their own progress and tasks. Teachers need to be given appropriate support to develop their confidence and competence in assessment practices and use a range of evidence to inform their decisions on student learning.

6. References

Aktamış, H., & Yenice, N. (2010). Determination of the science process skills and critical thinking skill levels. Procedia - Social and Behavioral Sciences, 2(2), 3282-3288.

Barton, Calabrese, Angela,. (1998). *Feminist Science Education*. Columbia University Press: Teachers College Press.

Baumfield, Vivienne. (2006). Tools for pedagogical inquiry: the impact of teaching thinking skills on teachers *Oxford Review of Education, 32*(2), 185-196.

Bernstein, R. (1983). *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*. Pennsylvania: Penn Press.

Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education, 26(3-4)*, 407-419.

Bybee, R. W. (2009). *The BSCS 5E instructional model and 21st century skills*. [Available on-line: http://itsisu.concord.org/share/Bybee_21st_Century_Paper.pdf] A commissioned paper prepared for a workshop on exploring the intersection of science education and the development of 21st century skills.

Csapó, B. (2010): Goals of learning and the organization of knowledge. In E. Klieme, D. Leutner & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik* (pp. 12-27). Weinheim, Germany: Beltz.

Csapó, B. (2012). Developing a framework for diagnostic assessment of early science. In S Bernholt, K. Neumann, & P. Nentwig (Eds.). *Making it tangible – Learning outcomes in science education,* (pp. 55-78). Münster: Waxmann.

Csíkos, C., & Csapó, B. (2011). Diagnostic assessment frameworks for mathematics: Theoretical background and practical issues. In B. Csapó & G. Szabó (Eds.). *Framework for diagnostic assessment of science* , (pp. 137-162) .Budapest: Nemzeti Tankönyvkiadó.

Csíkos, C., & Verschaffel, L. (2011). Mathematical literacy and the application of mathematical knowledge. In B. Csapó & M. Szendrei (Eds.), *Framework for diagnostic assessment of mathematics* (pp. 57-93). Budapest: Nemzeti Tankönyvkiadó.

Eklöf, A. (2013). A long and winding path: Requirements for critical thinking in project work. *Learning, Culture and Social Interaction*, 2(2), 61-74. doi: dx.doi.org/10.1016/j.lcsi.2012.11.001

Ennis, R. H. (1995). *Critical thinking.* New York: Prentice Hall.

Fradd, S.H., Lee, O., Sutman, F.X., & Saxton, M.K. (2001). Promoting science literacy with English language learners through instructional materials development: A case study. *Billingual Research Journal, 25 (4),* 417-439.

Harlen, W. (2005). *Teaching, learning and assessing science 5-12*. Sage Publications.

Heritage, M. (2013). Gathering evidence of student understanding. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 179-195). London: SAGE

Korom, E., B. Németh, M., Nagy L., & Csapó, B. (2012). Diagnostic assessment frameworks for science: Theoretical background and practical issues. In B. Csapó & G. Szabó (Eds.). *Framework for diagnostic assessment of science* , (pp. 147-174) .Budapest: Nemzeti Tankönyvkiadó.

Kirk, R.E. (1995) *Experimental Design: Procedures for the Behavioral Sciences* (3rd edition.). Pacific Grove, CA: Brooks/Cole.

Krajcik, J, Blumenfeld, Phyllis , Marx, Ron , & Soloway, Elliot. (2000). Instructional, Curricular, and Technological Supports for Inquiry in Science Classrooms. In Jim Minstrell & Emily H. van Zee (Eds.), *Inquiring into Inquiry Learning and Teaching in Science*. Washington: American Association for the Advancement of Science.

Matthews, B. (2006). *Engaging Education. Developing Emotional Literacy, Equity and Co-education*. Buckingham: McGraw-Hill/Open University Press

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking.* Pacific Grove, CA: Midwest Publications Critical Thinking Press.

OECD (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD. [Retrieved from: http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33693997.pdf]

OECD (2013). PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Overton, W. F. (Ed.). (1990). *Reasoning, necessity, and logic: Developmental perspectives*. Psychology Press.

Thomas, L. & Wareing, S. (1999). Language, society and power. London: Routledge

Topping, K. (2013). Peers as source of formative and sumamative assessment. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 395-412). London: SAGE.

Walker, M. (2007). *Teaching inquiry based science*. LaVergne, TN: Lightning Source.

Wenning, C. J. (2007). Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Teacher Education Online*, *4(2)*, 21-24.

7. Appendix

Overview of the implementation of SAILS inquiry and Assessment Unit

Unit Title	Overview of Unit implementation
Acids, bases, salts	This unit aids students to explore acids, bases and salts as substances that are used in everyday life. The inquiry approach used in all the case studies was that of guided inquiry, where the teacher identifies the problem and poses multiple questions that lead the students to answer inquiry questions. In this mode of inquiry, students are able to exploit pre-existing knowledge in order to formulate initial hypotheses, which will then help them structure their research (planning investigations). The inquiry skills of developing hypotheses, forming coherent arguments, planning investigations, and working collaboratively were assessed in different ways. Additionally the content knowledge and evidence of scientific literacy and scientific reasoning was assessed. The assessment methods used include teacher observation, classroom dialogue, assessment of student artefacts and student self-assessment. For some skills, the assessment was carried out after class and was based on a written artefact produced in class. In other situations, formative assessment guided the student learning during the class.
Collision of an egg	This unit asks students to solve an unstructured problem in the theme of mechanics – “what factors influence forces during collision?” To understand the interactions during a collision, the students study the impacts on an egg. Two approaches are recommended, first to consider “what factors make it possible for the egg to land safely?” and secondly “From how high can you drop an egg into a bucket of flour, without it breaking?” Through this activity, students explore the connection between force and momentum and can apply this knowledge in the context of road safety. This unit focuses on the inquiry skill planning investigations (designing an experiment), in particular considering variables. In addition, students engage in developing hypotheses, and their motivation can be enhanced through immersion in doing science. Working collaboratively with peers is important when developing and implementing the research plan. The teaching approach was open or open/guided inquiry in all cases; students were free to plan the experiment but the materials and equipment were provided. Inquiry skills assessed were planning investigations, developing hypotheses and working collaboratively. Possible assessment opportunities include teacher observation, evaluation of student artefacts using rubrics and self-assessment.
Electricity	This unit provides an introduction to electrical conductivity and electric circuits and is recommended to be implemented after students have studied electrostatics. Three activities are presented and use a guided inquiry-based approach with students at lower second level. Activity A introduces the students to the topic through a whole class brainstorming activity, and students construct a mind map of the topic based on their prior knowledge. In Activity B, the students design and assemble a simple working electric circuit. Students then use their circuits for planning and carrying out an investigation on the conductivity of every-day objects and materials (Activity C). This unit presents opportunities for assessment of several inquiry skills, in particular planning investigations and working collaboratively, as well as improving students’ scientific reasoning capabilities and scientific literacy. The

	assessment methods described in the unit include teacher observation, group brainstorming and use of student artefacts.
Light	Students examine the physical properties of light and its interaction with materials in a predominately qualitative fashion. A series of eight activities are described that aim to develop students' understanding of the concept of light and its characteristics. The unit activities are presented as a guided inquiry-based approach and an individual student worksheet is provided for each activity. This unit presents several opportunities for the assessment of different inquiry skills, and in particular, planning investigations, developing hypotheses, forming coherent arguments and working collaboratively. In addition, students can develop their scientific reasoning and scientific literacy skills. The assessment methods used across the activities of the unit include teacher observation, classroom dialogue, student worksheets and self-assessment.
Natural selection	This unit focuses on natural selection and the Darwinian theory of evolution, which is part of the biology curriculum at upper second level in most European countries. The topic is addressed in a structured, hands-on activity, during which students simulate a gene pool and the random selection of alleles. In all cases, the teaching approach was guided inquiry, although teachers also allowed open inquiry where feasible. The inquiry skills developed in this unit include planning investigations, forming coherent arguments and working collaboratively. Skills in scientific reasoning, such as collecting data, drawing conclusions are enhanced and students' scientific literacy is enriched through comparisons between the physical simulation and the real world. This activity is recommended for implementation at upper second level, where students have sufficient mathematical knowledge to numerically analyse a large quantity of data, and have conceptual understanding of the biology involved. The assessment methods described in the unit include teacher observation, use of student artefacts and classroom dialogue. Skills assessed included planning investigations, working collaboratively, scientific reasoning and scientific literacy.
Reaction rates	This unit uses effervescent vitamin C tablets to introduce students to the concepts of gas production in the reaction of acid with carbonate, and rates of reaction and factors influencing reaction rate. Three main activities aimed at lower second level are outlined, although they can be further extended and adapted for upper second level. The activities can be carried out in a sequence of lessons, which would require about ten class periods, or a specific activity can be targeted, requiring about two class periods depending on the skills to be assessed. Students can develop a number of inquiry skills, in particular planning investigations and working collaboratively. They furthermore have the chance to progress their scientific reasoning capabilities and scientific literacy, through critiquing experimental design, interpreting and analysing data and graphical interpretation, and thus develop skills in forming coherent arguments. The teaching approach in all case studies was that of an open/guided inquiry. The inquiry skills assessed were planning investigations, and working collaboratively, as well as the assessment of scientific reasoning (drawing conclusions). A broad range of assessment methods was utilised, ranging from in-class observation to evaluation of artefacts after the lessons, and including peer- and self-assessment.
Speed	Two activities are presented here for introducing the concept of velocity. Kinematics is a topic found on both lower and upper level science curricula across

	Europe, and forms the basis for many advanced topics in physics. The activities are presented as a bounded inquiry and each activity is expected to take one 45-minute lesson. This unit can be used for development of many inquiry skills, such as planning investigations (assessed in all case studies), developing hypotheses, forming coherent arguments and working collaboratively. In addition, students develop their scientific reasoning and scientific literacy. Possible assessment opportunities include teacher observation and classroom dialogue, evaluation of student artefacts and self-assessment.
Proof of the pudding	This unit outlines an inquiry activity in which the students (plan to) prepare a "good" pudding. This can focus on biological aspects – nutrition, energy content of foods, quality of nutrients, healthy lifestyles – and chemical concepts – groups of organic compounds, colloid systems, and sol gels. The close connection with everyday life and learning based on hands-on activities raise the students' interest. The three activities first introduce the topic, develop into planning and implementing an investigation and end with reflection on new knowledge. The teaching approach in the case studies was generally that of guided inquiry (open inquiry in one class). Through this activity, students develop their inquiry skills in planning investigations by distinguishing alternatives and constructing models, as well as skills in developing hypotheses, forming coherent arguments – setting variables, handling quantities, making comparisons, making judgements and decisions, analysing and critiquing experiments – and working collaboratively. The assessment opportunities described include student observation, group discussion or presentation and evaluation of student artefacts.
Black tide: oil in the water	This unit focuses on the study of the effects of an oil spill on our coast. Students investigate oil spills using a model system to simulate the behaviour of oil in water and identify factors that influence the spread of oil. Students can consider the ecological impact of an oil spill, and the challenges that are posed to scientists and society. This unit is recommended for implementation at lower second level, as a bounded or guided inquiry. This unit can be used for development of many inquiry skills, in particular planning investigations, developing hypotheses and working collaboratively. In addition, students can develop their scientific reasoning skills through collecting data and drawing conclusions, and enrich their scientific literacy by critically evaluating their investigations. Planning investigations was assessed in four of the case studies, while developing hypotheses and working collaboratively were also widely assessed.Assessment methods include teacher observation, student artefacts and peer- and self-assessment.
Floating Orange	This unit focuses on studying floating oranges as a model system to relate the physics concept of density and Archimedes principle with students' daily lives. This unit was designed as an inquiry activity that allows teacher to assess during the process of the inquiry. Students work in groups to develop hypotheses about the behaviour of oranges in water, and verify their hypotheses by experimentation. This unit is recommended for implementation at lower second level and the unit activities are presented as an open inquiry. Implementation of this unit is suggested for the assessment of students' skills in developing hypotheses and planning investigations, as well as enhancing their scientific literacy as they learn to explain the science behind the observed phenomena. Planning investigations and developing hypotheses were assessed in most cases, while working collaboratively was assessed in four of the eight case

	studies. Key assessment methods used include classroom dialogue, teacher observation and evaluation of student artefacts.
Food labels	This unit was designed to aid students to understand food labels and the composition of foods. Through the four outlined inquiry activities, students learn to look at the composition of foods and the amounts needed to keep someone healthy. In this way, students become equipped with sufficient knowledge and skills to make the choices that they need to when it comes to their own diet. The unit is recommended for implementation as a guided inquiry at lower second level. Two key skills are identified for development in this unit. Scientific reasoning, in particular proportional reasoning, is developed as students compare different amounts and types of food in their diet. Students' skills in working collaboratively are also developed, through discussion and teamwork. In all cases the unit was implemented as a guided inquiry, with some open opportunities. In addition to the assessment of scientific reasoning and working collaboratively, opportunities for the assessment of skills in developing hypotheses, planning investigations and forming coherent arguments were identified. The assessment methods described include classroom dialogue, teacher observation and evaluation of student artefacts.
Global warming	This unit aims to enable students to consider scientific data and determine whether or not the evidence supports the phenomenon of global warming. An additional activity presents an opinion piece, which the students should critique to judge its scientific merit. This activity may be implemented at lower or upper second levels depending on the curriculum's objectives, and is designed as a bounded inquiry. The key skills for development through these tasks are forming coherent arguments, working collaboratively and scientific reasoning. Students also enrich their scientific literacy through evaluation and use of scientific information. The assessment method emphasised for use is that of self-assessment, and rubrics are provided for students to use for evaluation of their work.The key skills assessed were forming coherent arguments, scientific reasoning and scientific literacy, with an emphasis on analysis and interpretation of scientific data and distinguishing opinions from facts. The assessment methods used include self-assessment, peer-assessment, classroom dialogue and evaluation of student artefacts.
Household v natural environment	This unit focuses on the environmental implications of the use of cleaning agents. Students investigate the growth of cress in various conditions, allowing them to determine the impact of commonly used household chemicals on the environment. Students assess the consequences of daily decisions taken in their homes and thus develop a sense of responsibility for the actions they take. This unit is recommended for implementation at both lower and upper second level, as a guided or open inquiry conducted over two lesson periods. This unit can be used for development of many inquiry skills, in particular planning investigations, developing hypotheses and working collaboratively. In addition, students can develop their scientific reasoning skills through collecting data and drawing conclusions, and enrich their scientific literacy by critically evaluating their investigations. The main skills assessed were planning investigations, working collaboratively and forming coherent arguments. This activity was shown to enrich students' scientific literacy, in particular the ability to present scientific data and to understand the environment impact of household chemicals. The assessment was based on teacher observation and the

	evaluation of students' presentations.
Plant nutrition	This unit aids students to learn about photosynthesis, a topic that features in curricula for second level education across Europe. In this unit, students use algae immobilised in jelly balls (alginate) to acquire evidence that light is necessary for photosynthesis to occur. These activities help them to connect observed phenomena and scientific theory. For lower second level, it is appropriate to implement a guided inquiry, although application of bounded inquiry at upper level could be considered. Students are provided the opportunity to develop inquiry skills such as planning investigations (planning and rationale, data recording, graphical representation), developing hypotheses, forming coherent arguments (reasoning and argumentation) and working collaboratively (discussing their decisions and conclusions). The teaching approach used in all case studies was open/guided inquiry. Inquiry skills assessed were planning investigations, developing hypotheses, forming coherent arguments, working collaboratively and scientific reasoning. Several assessment methods are described, including classroom dialogue, teacher observation and evaluation of worksheets, presentations or other student artefacts.
Polymers	This unit focuses on studying properties of plastic materials (density, thermal and electrical conductivity, combustibility) through experimentation. Students develop hypotheses about expected properties based on their previous knowledge and verify them subsequently by experimentation. This unit is recommended for implementation at upper second level and the unit activities are presented as a guided inquiry. Activity A introduces the determination of density of plastic materials by comparing with water density, while Activity B looks at combustion properties of plastic materials. Further activities look at their thermal stability and thermal conductivity (Activity C) and electrical conductivity (Activity D). This unit can be used for development of many inquiry skills, in particular developing hypotheses and planning investigations. Working collaboratively and planning investigations were assessed in most case studies, while the assessment of developing hypotheses, forming coherent arguments and scientific reasoning is also reported. The assessment methods described include teacher observation, use of student artefacts and self-assessment.
Up there… how is it?	This unit aids students to learn about the concept of gravity and offers an opportunity to learn about the International Space Station (ISS); understanding its impact on the scientific, technological development and society. In this unit, students are encouraged to develop interest and curiosity about space exploration. While recommended for upper level physics students, these activities could be explored with different disciplinary areas, namely chemistry, biology and geology, or adapted for implementation at lower second level. The teaching approach in all case studies was that of an open/guided inquiry. Through these activities, students are provided the opportunity to develop inquiry skills such as planning Investigations, developing hypotheses and working collaboratively, as well as progressing their scientific literacy and scientific reasoning capabilities. Inquiry skills assessed were planning investigations, reasoning (observation skills) and forming coherent arguments. Possible assessment opportunities include student observation, group discussion or presentation and evaluation of student artefacts.
UV radiation	Four activities are presented in this unit for introducing the concept of UV radiation. In particular, this unit addresses sources of UV radiation, potential

	health and safety considerations and methods of detection. These investigations are achieved using UV reactive beads (or a UV sensor). This is suitable for implementation at lower level (age 12-16). The unit activities are presented as an open/guided inquiry approach and when implemented in full is expected to require 3 hours (180 minutes). This unit can be used for development of many inquiry skills, in particular planning investigations. In addition, students can develop their skills in developing hypotheses and forming coherent arguments, and enhance their scientific reasoning and scientific literacy. The teaching approach was open/guided inquiry or bounded inquiry. Developing hypotheses and planning investigations were assessed in all case studies; the assessment of working collaboratively and scientific reasoning is also reported. Possible assessment opportunities include teacher observation, student artefacts, use of rubrics and self-assessment.
Which is the best fuel	This unit aims to encourage students to realise that fuels have different heats of combustion and allow them to realise that the meaning of "best" can change depending on the context. This is achieved by planning and carrying out an experiment to measure heat energy changes and finding enthalpies of combustion experimentally. This activity may be implemented at lower or upper second levels depending on the curriculum's objectives and full implementation requires four lessons. Through this unit, students are provided the opportunity to develop a number of inquiry skills such as developing hypotheses, planning investigations (design and conduct an experiment), and forming coherent arguments (draw appropriate conclusions using reasoned arguments). In addition they build their scientific reasoning capabilities by collecting meaningful data, and enrich their scientific literacy through analysis of scientific data and presentation of scientific conclusions. The teaching approach used in all case studies was bounded or guided inquiry, with some open opportunities. All four SAILS inquiry skills were assessed – planning investigations, developing hypotheses, forming coherent arguments and working collaboratively – as well as scientific reasoning capabilities. The assessment methods described include classroom dialogue, teacher observation, group discussion or presentation and evaluation of student artefacts.
Woodlice	This unit outlines an activity that is intended to aid students in learning about the environment, ecology, and animal behaviour. Students investigate the living conditions of woodlice, which are common in large parts of Europe and are easy to handle. The expected learning outcomes are: (1) learn to plan, perform and evaluate an experimental study, and (2) identify and explain ecological relationships using scientific concepts, models, and theories. These learning outcomes are part of the science curriculum at lower second level across Europe. Throughout the activities students will have opportunities to practice a range of other inquiry skills, such as collecting and interpreting data (planning investigations), drawing appropriate conclusions (forming coherent arguments), and reporting and discussing results (scientific reasoning). Suggested assessment tools are provided in the unit, but it was the teachers' choice to select what inquiry skills to develop and assess. The teaching approach varied across the four case studies for example one teacher formed a guided inquiry, while another engages the students in open inquiry. In all case studies developing hypotheses was assessed, while the assessment of other skills varied across the case studies.