Report from the FP7 project:

Assess Inquiry in Science, Technology and Mathematics Education

Report of outcomes of the expert workshop on formative assessment in STM and IBE

Silke Rönnebeck, Mathias Ropohl, & Olaf Köller
with the assistance of Hilda Scheuermann

Delivery date	15.10.2013
Deliverable number	D 2.6
Lead participant	Leibniz-Institute for Science and Mathematics Education (IPN), Kiel, Germany
Contact person	Silke Rönnebeck (roennebeck@ipn.uni-kiel.de)
Dissemination level	PU

Table of Contents

Summary

As part of its work reviewing the current state of the art of formative and summative assessment, WP2 organized and conducted an expert workshop on the use of formative assessment in inquiry-based education (IBE) in science, technology and mathematics (STM). The workshop was held in Berlin on August 23 and 24 and was attended by representatives from all ASSIST-ME partner countries. In addition to interactive working sessions, three international experts presented insights into their work within the field of formative assessment in science and mathematics. The aim of the workshop was to discuss current research in the field of formative assessment of IBE in STM and to relate these findings to the objectives of the ASSIST-ME project, thus providing a basis for the future work. The workshop resulted in a list of key characteristics for the assessment methods that are going to be developed within the project as well as a list of possible constraints that have to be taken in account during the development process.

1. Introduction

As part of its work, WP2 organized and conducted an expert workshop on the use of formative assessment in inquiry-based education (IBE) in science, technology and mathematics (STM). The workshop was held in Berlin on August 23 and 24. It was attended by representatives from all ASSIST-ME partner countries (see Table 1). In addition to interactive working sessions, three international experts presented insights into their work within the field of formative assessment in science and mathematics.

Table 1: List of workshop participants

Name	Institution	Country
Rose Clesham	Pearson Education International	UK
Costas Constantinou	University of Cyprus	CY
Jens Dolin	University of Copenhagen	DK
Bob Evans	University of Copenhagen	DK
Erin Marie Furtak	University of Colorado	USA
Regula Grob	University of Applied Sciences and Arts Northwestern Switzerland	CH
Jan-Christoph Hadenfeldt	Leibniz Institute for Science and Mathematics Education	D
Birgit Harks	German Institute for International Educational Research	D
Christine Harrison	King's College London	UK
Monika Holmeier	University of Applied Sciences and Arts Northwestern Switzerland	CH
Eckhard Klieme	German Institute for International Educational Research	D
Olaf Köller	Leibniz Institute for Science and Mathematics Education	D
Florence Le Hebel	The National Center for Scientific Research (CNRS)	F
Jan Alexis Nielsen	University of Copenhagen	DK
Pasi Nieminen	University of Jyväskylä	FI
Nicos Papadouris	University of Cyprus	CY
Lukáš Rokos	University of South Bohemia	CZ
Silke Rönnebeck	Leibniz Institute for Science and Mathematics Education	D
Mathias Ropohl	Leibniz Institute for Science and Mathematics Education	D
Hilda Scheuermann	Leibniz Institute for Science and Mathematics Education	D
Carl Winsløw	University of Copenhagen	DK
Iva Žlábková	University of South Bohemia	CZ

The aim of the workshop was to discuss the current state-of-the-art of formative assessment of IBE competences in STM to develop ideas and recommendations for the future work of the ASSIST-ME project .

ASSISTME

2. The expert presentations

Three international experts in the field of formative assessment were invited to present their work and engage in discussions about the implications of their work for the future work in ASSIST-ME.

2.1 Formative assessment: general concepts, recent debates in Germany, and findings from experimental studies in mathematics education

The first presentation was given by Eckhard Klieme from the German Institute for International Educational Research (DIPF) who together with his colleague, Birgit Harks, talked about 'Formative assessment: general concepts, recent debates in Germany, and findings from experimental studies in mathematics education'. The talk focused on feedback as one characteristic property of formative assessment. Different types of formative assessment were distinguished based on their position on a continuum ranging from 'interactive (instructional) feedback' (on-the fly) to 'measurement' (benchmark tests that are called formative although they are mostly summative in nature). The first half of the presentation dealt with questions concerning the effectiveness of feedback and its role within the German educational system. The effectiveness of feedback can be influenced by different factors like e.g. the frame of reference, the task/process relation or students' attributions – beneficial feedback characteristics often conflict with grading. Within formative assessment cycles, feedback constitutes the link between assessment and the adaptation of teaching. Although the formative assessment cycle includes all the concepts that guide recent education in Germany, the notion of assessment still seems very unpopular with German teachers. They prefer the notion 'individual support cycle' since the idea that assessment can support learning still seems alien to them. In the second part of the presentation, empirical results from the project 'Conditions and Consequences of Classroom Assessment (Co^2CA)' were presented. One question investigated within the project was e.g. which types of assessment and feedback teachers generally apply in their everyday mathematics instruction. Three types were identified, namely grading, verbal feedback and student-centered activities like peer/self-assessment or portfolios.

The first part of the discussion focused on the challenge to ensure that really the phenomenon is studied and not just a special framework. It was widely acknowledged that a major problem of research in formative assessment and feedback is that, even in reviews, so many different notions exist of what constitutes these approaches. To meet this challenge, one has to make very explicit definitions and position oneself on the continuum from 'interactive (formative) assessments (e.g. on-the-fly)' to 'measurement (e.g. benchmark tests)'. For many people (researchers as well as teachers) on-the-fly would not even qualify as assessment. Moreover, it might be called differently in different countries or educational contexts. In the literature, there is some disagreement whether summative assessments could (or should) be used formatively or vice versa (see ASSIST-ME deliverable D2.4). Within ASSIST-ME, however, the assessment ac-

tivities itself are important and can be used either in a summative or formative way. These activities share three characteristics: they provide information, they include feedback and they allow teachers to adapt their instruction. Even on-the-fly assessments can be used as summative assessments, e.g. Danish teachers often grade their students based on on-the-fly interactions. Despite the great importance of this oral-based grading, however, there seems to be almost no research on this, at least in German-speaking countries.

Another issue in the discussion dealt with the separation of feedback and adaptive teaching. Whereas in practice they are often combined, in theory just providing feedback does not necessarily tell teachers and students what to do next. Emphasizing that it is a two-step process might make the process more understandable for teachers. A last issue, eventually, focused on problems of equity since rather strong indicators exist that some forms of formative assessment have problems with equity (e.g., gender, language proficiency, etc.). Teacher bias, however, was seen as a problem in all forms of assessment and not as characteristic for formative assessment. It can be found in verbal interaction as well as in grading. One means to reduce bias has been shown to be good classroom management.

2.2 Formative Assessment from an International (US) Perspective

In the second talk 'Formative Assessment from an International (US) Perspective', Erin Furtak from the University of Colorado in Boulder presented the recent situation of formative assessment in the US. She distinguished three areas: trends in formative assessment policy, trends in formative assessment research and remaining challenges in formative assessment practice. Within the policy area, these trends included the development of commercially available 'standardized formative assessments' (that schools or states can purchase and that should provide teachers with diagnostic information) and observation protocols to track teacher performance (in which one part relates to their formative assessment practice). In the field of research it can be observed that there still is an on-going discussion about the definition of formative assessment. As one possible solution, a categorization scheme was presented that distinguishes three broad aspects of formative assessment – providing information, taking actions based on this information and achieving a desired effect based on these actions – and that categorizes the definitions based on the aspect they are focusing on. Another aspect from the field of research deals with the domain-dependency of formative assessment. Observing teachers in discussions about formative assessment, one finds that they talk a lot about the science involved (subject-dependent) and a lot about teaching (subject-independent). A connection between these two fields is achieved via talking about 'science teaching'. Concerning remaining challenges in formative assessment practice, research shows that this practice continues to be challenging for teachers to realize and that better models for professional development that focus on individual teacher development are needed.

In the discussion it was asked whether teachers included criteria for formative assessment in their conversations which they did although they didn't necessarily explicitly call

them criteria. Although the research presented had no special focus on inquiry, epistemic aspects of IBE like e.g. developing questions, explaining or reflecting were regarded as playing an important role in formative assessment with the teacher guiding the students in this process.

2.3 Assessment for Learning in the UK

In her presentation 'Assessment for Learning in the UK', Christine Harrison from King's College London who is leading WP7 in ASSIST-ME, talked about Assessment for Learning (AfL) as the UK approach to improve classroom assessment. Her talk focused on the underlying principles of AfL, its conceptualization in the classroom and problems encountered in its widespread implementation. One main purpose of AfL is seen in the development of self-regulated learning. This is also emphasized in the definition of AfL:

> *'[AfL] is classroom assessment which focuses on the learning as it is taking place and its function is to bring about improvement. Both teachers and learners need to be involved but ultimately it is the learner who has to take action'.*

Whereas the learner has to change his role from passive to involved, the teacher has to become a conductor rather than a controller. It has been found that teachers often have difficulties in perceiving the changes they need to make in their day-to-day practice in order to change their instruction towards AfL. One way to overcome these problems is seen in collaborative action research because it 'intertwines professional development and research in an iterative process that attempts to recognise, document, and monitor change both as it happens and through reflective review'. In the second half of her presentation, Chris Harrison talked about experiences from SAILS – Strategies for Assessment of Inquiry Learning in Science. SAILS is another EU-FP7 project dealing with the assessment of scientific inquiry that started one and a half years earlier than ASSIST-ME. Within SAILS, it has been found that getting teachers to think about (formative) assessment brings them to include more inquiry into their teaching. Teachers have started to make room for inquiry within the regular curriculum and the inquiry has become more open. Nevertheless, problems with teacher assessment have also become evident. These problems are on the one hand related to teachers worrying about assessment (or test) fairness like e.g. the perceived conflict between collaborative work and individual performance or the fact that during authentic activities it is almost impossible to collect data about every single student. On the other hand teachers often express a lack of confidence not only in their own inquiry and assessment skills but also a general lack of public confidence in teacher assessments.

The discussion focused on rather practical questions. Since it is quite complex and challenging for teachers to do inquiry and assessment at the same time, it was recommended that the project should first go for teachers who have experience in inquiry teaching. Later, when the methods have been established, one might try to transfer them to less experienced teachers. Moreover, it was affirmed that collaborative learning is a prerequisite for AfL if the definition of AfL includes the component of self-regulated learning. Another issue concerned the assessment of individual performance in collaborative learning. Here one may benefit from expertise in other subjects like e.g.

drama where this assessment works well. One might involve drama teachers in teacher training activities where they might show how individual assessment in collaborative learning works in drama classes and how it might be transferred to STM. Teacher training was also the focus of the last issue which dealt with the question if teachers can be supported in adopting AfL by classroom videos. In France it was found that teachers had difficulties noticing especially on-the-fly assessment when watching videos. This seemed to be a general phenomenon. Instead of focusing on the learning, teachers tend to focus on criticizing and saying what they would do differently. To avoid this, videos should focus on the learners (the learning) instead of on the teachers (the teaching).

3. The Workshop session

Following a presentation of the main findings of ASSIST-ME deliverable D2.4 'Report on current state of the art in formative and summative assessment in IBE in STM', the workshop participants worked in small groups to try and connect different aspects of inquiry that had been identified during the literature review with different assessment methods (see Figure 1).

Aspects of inquiry	Links	Formative assessment methods
Diagnosing problems		Concept maps
Critiquing experiments		Mind maps
Distinguishing alternatives		Portfolios
Planning investigations		Science notebooks
Researching conjectures		Effective questioning
Searching for information	???	Assessment conversations
Constructing models		Accountable talk
Debating with peers		Quizzes
Forming coherent arguments		Multiple choice
…		Constructed response
		Open ended
		…

Figure 1: Links between aspects of inquiry and formative assessment methods.

Key questions that should be addressed were whether aspects of inquiry exist that should be preferably assessed within ASSIST-ME (and, if yes, by which method(s)) or if certain assessment methods are particularly suited for assessing certain aspects of IBE. Following the group discussions, the groups presented their results in the form of short poster presentations.

The presentations showed that the groups used quite different approaches to the task. All groups, however, reported that they felt somehow overwhelmed especially by the number of inquiry aspects which resulted in different attempts of grouping them e.g. according to their relatedness, their degree of specialization vs. generalization or their frequency of occurrence in daily teaching practice. Selected results from the group presentations and following discussions are:

- Aspects of inquiry
  - The aspects are different in nature (e.g. skills vs. activities)
  - The aspects are too general, it is necessary to define exactly what is meant by them
- Assessment methods
  - Whereas some methods may be related more closely to certain groups of aspects, others might be considered more general or overarching methods (like e.g. portfolios, notebooks, questioning, learn logs or constructed/open response items)
  - To use portfolios in a formative way means to understand them iteratively, not as a final product

- Diagnostic sheets have the potential to assess quite different aspects of inquiry e.g. the modeling process in mathematics (see Klieme's presentation)
- Technology might help in providing feedback to students and teachers (e.g. the Socrative-Website or Apps); Socrative e.g. is a free internet resource that is accessible by smartphones. Students e.g. type in their hypotheses and they instantly appear anonymously on screen for everyone to see. Students can then vote which hypothesis they consider to be the most sensible or reasonable and the hypotheses are re-arranged according to the votes. This allows for student self-assessment of how the own idea is perceived in class and for teachers getting a good idea what the whole class is thinking.
- Open-ended problems, e.g., have the potential for assessing different aspects of inquiry like e.g. planning investigations, searching for alternatives/modifying design, dealing with uncertainty and diagnosing problems/identifying questions
- A strong contingency is seen on content area and institutional conditions

One group took a more theoretical approach and re-formulated the key questions (see Figure 2) to:
1. What are we assessing?
2. How are we assessing?
3. What instruments/tools do we need?

'What are we assessing' addresses the question whether assessment tools should be able to assess very subject-specific or more or less subject-independent competences. The recommendation was to focus on the latter, i.e. broad, cross-cutting competences like e.g. modeling, planning & carrying out investigations or constructing evidence-based arguments, and think about types of classroom processes that we want to assess. In this context, it was stressed – similar to the recommendations from the other groups – that the aspects of inquiry identified in the literature review should, at the current status of the project, be grouped into broader competences. It was felt that at this stage of the project there are too many.

The second question expresses a need for clear and decisive definitions. It needs to be defined what is meant by methods, formats and tools and what these definitions require or what their consequences are, respectively. With respect to methods, one needs to think about e.g. process information (e.g. to collect data), what features are expected (e.g. reliability, validity) and how these features can be transformed to make them meaningful for everyone involved in the project including the teachers. Once the data is collected and the evidence is there, one also needs to think about analytical procedures. Who is going to do the analysis? Will the teachers be expected to do the structuring, analysis, evaluation and communication in a meaningful way completely on their own or who else (and in which way) should be involved in the analytical process? In this context, it was stressed that communication should be an integral part of the methodology.

In due time, the project should decide on sets of assessment methods – this decision, however, should not only be about formats but also about competences, processes and analytical procedures. With respect to the last question, several issues came up. First of all, the answer to this question should be based on the type of classroom processes we want to collect data about. The project needs formats that allow for looking at outputs of what students are constructing but also for looking at learning processes. Moreover, the project should think about ways to support the teachers in their tasks and provide them with tools or instruments. Rubrics were regarded as an essential and indispensable tool in this respect. They can provide concise structures that allow teachers to organize information in a sensible way. Another important aspect is the development of the actual assessment tasks. In contrast to the competences that are supposed to be subject-independent, the assessment tasks should be content-oriented and embedded in the respective discipline. Moreover, the tasks should challenge the students and must allow for collection of meaningful data.
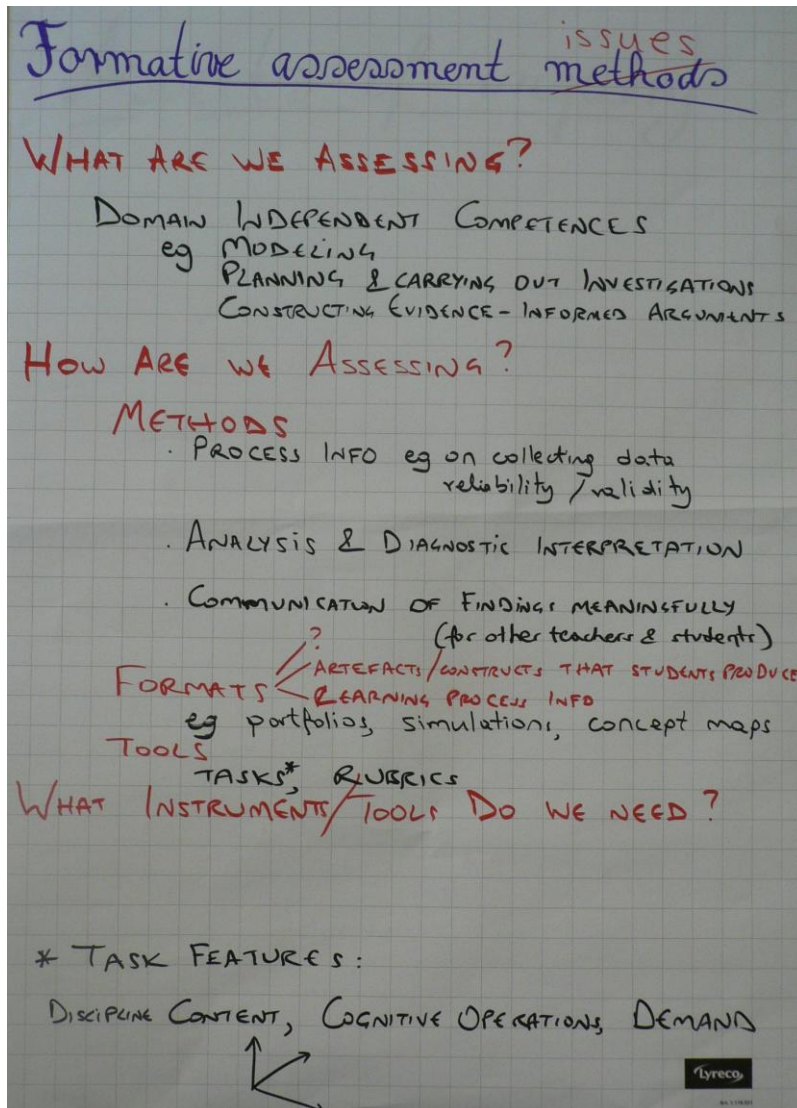

Figure 2: Outcome of one group discussion

4. Concluding Remarks and Outlook

The plenary discussion at the end of the workshop summarized the main outcomes of the workshop and their consequences for the further work of ASSIST-ME. It was agreed that

- ASSIST-ME should focus on subject-independent competences such as e.g. modeling, investigations or argumentation, and develop tools and mechanisms for assessing these competences.
- The assessment tasks themselves will be domain-rich and embedded in the different disciplines. They should aim at informing the implementation of whole learning progressions for developing the competences.
- A small number of assessment formats will be chosen and illustrative examples that are close to existing teaching practices will be developed
- The project should focus on teachers experienced in IBE (at least in the beginning). Training modules should include guidelines for effective teaching and learning.
- AfL needs to be embedded in specific pedagogical contexts. Hence, the pedagogical orientations need to be stated explicitly.

The workshop also ended up with some constraints or conditions the project needs to comply with. The assessment methods that are going to be developed

- should not be regarded as a lot of extra work by the teachers (time is a serious constraint for teachers) – and they should be tailored to most ordinary classrooms,
- should focus on the learning as it is taking place and provide ways to improvement,
- should involve both, teachers and learners – ultimately, however, it is the learner who has to take action,
- should include some ´modern'/IT-based methods (social media, cloud sourcing, …),
- should all be illustrated by an exemplary case and some guidelines for using them (to facilitate the later production of teacher training material)
- should have relevance in everyday life since this is an important attribute for the students.

In order to convince teachers to engage in the assessment process, the benefits of this engagement should be explicitly described in a way that is convincing for the teachers. With respect to specific assessment formats, the project should come up with criteria for promoting productive discourse and be able to capture the classroom dialogue. This explicitly includes giving good formative feedback on written assignments.

Figures

Tables