Report from the FP7 project:

Assess Inquiry in Science, Technology and Mathematics Education

ASSISTME

Recommendation report from D 2.1 – D 2.6 to be used as a foundation for WP 4 and 5

Silke Rönnebeck, Mathias Ropohl,
Sascha Bernholt, Olaf Köller, & Ilka Parchmann
with the assistance of Hilda Scheuermann & Sabrina Schütz

in collaboration with Sarah Hughes, Clare Green, & Vanessa Greene

Delivery date	15.10.2013
Deliverable number	D 2.7
Lead participant	Leibniz-Institute for Science and Mathematics Education (IPN), Kiel, Germany
Contact person	Silke Rönnebeck (roennebeck@ipn.uni-kiel.de) Mathias Ropohl (ropohl@ipn.uni-kiel.de)
Dissemination level	PU

Table of Contents

SUMMARY 4

1. INTRODUCTION 5

2. THEORETICAL BACKGROUND 7

2.1 Concept of competence 7

2.2 Definitions of IBE in STM 7
2.2.1 Science 8
2.2.2 Technology 8
2.2.3 Mathematics 9

2.3 Purposes of formative and summative assessment 9
2.3.1 Summative assessment 10
2.3.2 Formative assessment 10

3. SUMMARY OF THE LITERATURE REVIEW 12

3.1 Procedure of the literature review 12

3.2 State of the art of IBE 14

3.3 State of the art of formative and summative assessment 19
3.3.1 Assessment practices in science 19
3.3.2 Assessment practices in technology 22
3.3.3 Assessment practices in mathematics 22

3.4 Connections between IBE and assessment methods 23

3.5 State of the art of ICT 27
3.5.1 Theories and models in e-assessment 28
3.5.2 Evaluation of e-assessment models 28
3.5.3 Exemplars of e-assessment 29

3.6 State of the art in the ASSIST-ME partner countries 29

4. FINAL CONCLUSIONS AND RECOMMENDATIONS 33

4.1 Conclusions 33

4.2 Recommendations 35
4.2.1 What should be assessed? 35
4.2.2 How should it be assessed? 35
4.2.3 Which tools or instruments are needed for the assessment? 37
4.2.4 How can teachers be supported in the process? 39

REFERENCES **41**

FIGURES **49**

TABLES **49**

Summary

The EU-funded project 'Assess Inquiry in Science, Technology and Mathematics Education' (ASSIST-ME) investigates formative and summative assessment methods to support and improve inquiry-based approaches in European science, technology and mathematics (STM) education. To guide the design, development and implementation of assessment methods, in the first phase of the project a review of the state-of-the art in formative and summative assessment and inquiry-based education in science, technology and mathematics has been conducted by WP 2. The review included international research, national research within the ASSIST-ME partner countries, and definitions used by former and on-going EU-projects in the field of IBE. The results of this work have been published in the reports D 2.3 'National reports of partner countries reviewing research on formative and summative assessment in their countries', D 2.4 'Report on current state of the art in formative and summative assessment in IBE in STM – Part I and part II', D 2.5 'A definition of inquiry-based STM education and tools for measuring the degree of IBE' and D 2.6 'Report of outcomes of the expert workshop on formative assessment in STM and IBE'. Following a short introduction, this report summarizes the major outcomes of the previous reports in sections 2 and 3. The outcomes provide the basis for a set of conclusions and recommendations presented in section 4. These conclusions and recommendations reflect the work done and the insights gained during the first nine months of the ASSIST-ME project that are regarded as crucial for the future work within the project.

1. Introduction

One mayor objective of the ASSIST-ME project is the development of assessment methods suitable for enhancing inquiry-based education (IBE) of science, technology, and mathematics (STM) related competences. Based on these methods, strategies for formative and summative assessment of competences in STM that are adaptable to various European educational systems will then be identified (Dolin, 2012). Work package 2 (WP 2) in the ASSIST-ME project mainly focuses on synthesizing existing research on IBE as well as on formative and summative assessment of competences in STM (see Table 1). Among others, this synthesis includes a definition of inquiry-based education in STM (D 2.5), the collection and compilation of national reports of the partner countries on formative and summative assessment (D 2.3), and a report on the current state of the art in formative and summative assessment in IBE in STM (D 2.4).

Table 1: Deliverables from Work Package 2

Number	Title
D 2.1	Guidelines for partner countries to facilitate collecting national research on assessment (Rönnebeck & Köller, 2013a)
D 2.2	Synopsis of the literature review (Rönnebeck & Köller, 2013b)
D 2.3	National reports of partner countries reviewing research on formative and summative assessment in their countries (Rönnebeck, Bernholt, Ropohl, Köller, & Parchmann, 2013)
D 2.4	Report on current state of the art in formative and summative assessment in IBE in STM – Part I and part II (Bernholt, Rönnebeck, Ropohl, Köller, & Parchmann, 2013)
D 2.5	A definition of inquiry-based STM education and tools for measuring the degree of IBE (Ropohl, Rönnebeck, Bernholt, & Köller, 2013)
D 2.6	Report of outcomes of the expert workshop on formative assessment in STM and IBE (Rönnebeck, Ropohl, & Köller, 2013)

Details and first summaries of these diverse foci can be found in the respective reports. This report provides the next step in further summarizing the main messages of the diverse reports, to uncover interrelations and to derive recommendations for the further work within the ASSIST-ME project.

The major challenge for the literature review in WP 2 was that the field of interest is not clearly defined. In the subject areas of STM, there is still disagreement among researchers about what features define the instructional approach of inquiry-based teaching (Furtak, Shavelson, Shemwell, & Figueroa, 2012; Hmelo-Silver, Duncan, & Chinn, 2007), differing for example between narrow or expansive definitions of inquiry-based teaching. In some cases explicit definitions of some aspects of inquiry-based science teaching are given (e.g. innovative, activity-based, process oriented or discovery oriented), in others the meaning of the term remains less clear. Moreover, the above mentioned authors found that researchers and educators use a rich vocabulary to describe inquiry-based approaches e.g. in science education like inquiry-based teaching and learning, authentic inquiry, model-based inquiry, modelling and argumentation, project-based science, hands-on science, and constructivist science. Including technology and mathematics in the reflection makes the picture even more diverse. A similar problem occurs when researching literature on formative and summative assessment in the classroom (Black & Wiliam, 1998).

2. Theoretical background

The ASSIST-ME project combines three research fields: the concept of competence, inquiry-based education, and formative and summative assessment. The following chapters give an overview of important definitions.

2.1 Concept of competence

Within the ASSIST-ME project, a competence is understood as an ability that is based on a combination of skills, knowledge, and traits. It contributes to performances in particular domains but also in real-life situations and is thus closely related to the concept of scientific literacy (OECD, 2006). Weinert (2001) points out that these subject-specific competences can generally be learned depending on greater or lesser abilities in acquiring these competences. This category of learnable competences is mainly focused by the ASSIST-ME project.

Examples of domain-specific competences are e.g. to observe and describe natural phenomena accurately and to plan, perform, evaluate and reflect on experiments (science); to construct and produce technical tools and instruments e.g. to plan, manufacture, evaluate, and optimize technical solutions (technology); to develop and test hypotheses about functional relations that have been observed in the reality (from data) or in mathematical settings (mathematics).

In addition, an individual possesses key competences. These are transversal competences that can be successfully applied across a maximum number of different tasks (Weinert, 2001). Examples for transversal competences are e.g. to process and interpret data and results, to argue and communicate with peers and experts, and to develop a sense of responsibility and become a responsible-minded citizen (Dolin, 2012; Weinert, 2001).

Furthermore, Weinert (2001) distinguishes between competences and metacompetences. According to his definition, metacompetences are a combination of knowledge, skills, and/or strategies that are appropriate to organize and recognize available competences in adaptive and flexible ways (Weinert, 2001). Thus, he defines metacompetences as declarative or procedural knowledge about one's own competences. This third category of competences is not addressed by the ASSIST-ME project.

2.2 Definitions of IBE in STM

According to Anderson (2002) – whose definition formed the basis of the ASSIST-ME application (Dolin, 2012) – inquiry-based STM education includes students' involvement in questioning, reasoning, searching for relevant documents, observing, conjecturing, data gathering and interpreting, investigative practical work and collaborative discussions, and working with problems from and applicable to real-life contexts. Whereas these characteristics generally apply to all three subject areas, the ASSIST-ME application explicitly acknowledges that various meanings and forms of inquiry are possible in different disciplines and need to be addressed in the project. The following paragraphs are a short summary of D 2.5 and point out these subject-specific characteristics of IBE.

2.2.1 Science

Basically, one has to distinguish four different meanings of scientific inquiry summarized by Furtak et al. (2012): 1) scientific ways of knowing (i.e., the work that scientists do), 2) a way for students to learn science, 3) an instructional approach, and 4) curriculum materials. These four different meanings are often not distinguished. In these cases, it becomes not clear which perspective is meant. In addition, instead of 'inquiry' other terms and phrases are used, e. g. problem-based learning.

Different perspectives on scientific inquiry were also considered by Abd El Khalick et al. (2004) when suggesting four dimensions of IBE that reflect the complexity of this teaching and learning approach more specifically:

1) types of knowledge and understandings:
   conceptual, problem solving, social, and epistemic,
2) range of inquiry-related activities:
   e. g. problem-posing; designing investigations; collecting or accessing data; generating, testing, and refining models and explanations; communicating and negotiating assertions; reflecting; and extending questions and solutions,
3) range of skills:
   e. g. mathematical, linguistic, manipulative, cognitive and metacognitive skills
4) range of spheres:
   e. g. personal, social, cultural, and ethical

However, inquiry is a common teaching and learning approach as defined by Linn, Davis, and Bell (2004):

> *"[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."* (Linn et al., 2004, p. 4)

The EU-funded projects from the Seventh Framework Program (e.g. Mind the Gap, S-TEAM, FIBONACCI, and SAILS) use this definition as a common basis for discussion. For example, McLoughlin (2011) specified these aspects of IBE and defined further fundamental abilities for each aspect within the ESTABLISH project. Another example of such a more detailed description was published by Yiping and Blanchard (2007). Such a specification is imperative to the formative and summative assessment of IBE.

2.2.2 Technology

Teaching and learning approaches in technology education related to inquiry are problem- or project-based learning. In the first case, the starting point is the presentation of a technical problem. Students have to find an answer and consider alternative solutions (Fox-Turnbull, 2006). In the second case, the starting points are the presentation of a target setting and of materials that can be used to reach this target.

One major parallel to mathematics education is the aspect of problem solving. In technology education it is named engineering design that involves three main aspects: 'defining and delimiting an engineering problem', 'developing possible solutions', and 'optimizing the design solution' (National Research Council, 2012). In contrast to science and mathematics education, however, the physical product of the design process is the intentional objective. Central abilities are e. g. 'designing', 'modelling', 'testing', 'troubleshooting', 'observing', 'analysing', and 'investigating' (International Technology Education Association, 2007).

2.2.3 Mathematics

Teaching approaches or learning theories in mathematics that include characteristics of inquiry are problem-solving (Polya, 1945; 1957), problem-centred learning (Schoenfeld, 1985; 1992), inquiry mathematics (Cobb, Wood, Yackel, & McNeal, 1992; Cobb et al., 1991), and open approach lessons (Nohda, 1995; 2000). Furthermore, the Dutch approach of realistic mathematics education (Freudenthal, 1973) and the French theory of didactic situations (Brousseau & Balacheff, 1997) are mentioned in connection with inquiry in mathematics. Moreover, the Swiss concept of dialogic learning (Gallin, 2012) includes aspects of inquiry. Another relevant approach is problem-based learning (cf. Rocard et al., 2007).

Based on the characteristics of the mentioned approaches, the EU-funded FIBONACCI project explicitly describes IBE in mathematics:

> *„Like scientific inquiry, mathematical inquiry starts from a question or a problem, and answers are sought through observation and exploration; mental, material or virtual experiments are conducted; connections are made to questions offering interesting similarities with the one in hand and already answered; known mathematical techniques are brought into play and adapted when necessary. This inquiry process is led by, or leads to, hypothetical answers – often called conjectures – that are subject to validation."* (Artigue & Baptist, 2012, p. 4)

However, the difference to science education lays in the starting point of the inquiry process (Artigue, Dillon, Harlen, & Léna, 2012). In mathematics, problems are posed and proof as to whether a conjecture is true or false results from a logical demonstration. In science, phenomena and questions are considered, and models derive from an investigation including e. g. experiments and observations. This contrast shows that the main difference to science education lies in the "solution which is necessarily presented as a deduction from what was given in the problem to what was to be found or proved" (Schoenfeld & Kilpatrick, 2013).

2.3 Purposes of formative and summative assessment

Assessment is one of the most important drivers in education and a defining aspect of any educational system. It always involves the collection, interpretation and use of data for some purpose. The purpose and often also the way of data collection, however, may differ. In ASSIST-ME, we summarize these different purposes under the expressions of summative and formative assessment.

2.3.1 Summative assessment

Summative assessment has the purpose of summarizing and reporting learning at a particular time and for that reason is also called 'assessment *of* learning'. It involves processes of summing up by reviewing learning over a period of time or checking-up by testing learning at a particular time. Summative assessment has an undeniably strong impact on teaching methods and content (Harlen, 2007).

2.3.2 Formative assessment

Formative assessment, in contrast, has the purpose of assisting learning and for that reason is also called 'assessment *for* learning'. It involves processes of "seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002). Or, as Bell and Cowie (2001) put it: "[Formative assessment is] the process used by teachers and students to recognize and respond to student learning in order to enhance that learning, during the learning" (Bell & Cowie, 2001, p. 540). According to Black and Wiliam (1998) and Wiliam (2006), assessments are formative if, and only if, something is contingent on their outcome and the information is actually used to alter what would have happened in the absence of that information.

Klieme and Harks (2013) distinguish different types of formative assessment in view of its objectives and its location on a continuum reaching from informal to formal assessment settings (see Figure 1). For example, formative assessment 'on the fly' aims at interactive (instructional) feedback given to the students whereas embedded assessment aims at the measurement of students' competences.

Measurement

- benchmark tests (summative)
- embedded assessment
- check for understanding (CRESST: Power source)
- learning progression maps (Wilson: BEAR)
- curriculum-embedded assessment
- reflective lessons (Shavelson et al.)
- planned for interaction
- on the fly

Interactive (instructional) feedback

Figure 1: Different types of formative assessment (Klieme & Harks, 2013)

Feedback plays an important role in formative assessment. Sadler (1989) points out that feedback is defined as information about how successful something has been or how successfully it has been done. Furthermore, feedback is information about the gap between the actual performance level and the reference level. By the provision of information, three questions should be addressed by the student: Where am I going?, How am I going?, Where to go next? (Hattie & Timperley, 2007). Based on the feedback, the teaching and learning processes should be adapted to the actual performance level in order to support students individually (see Figure 2).

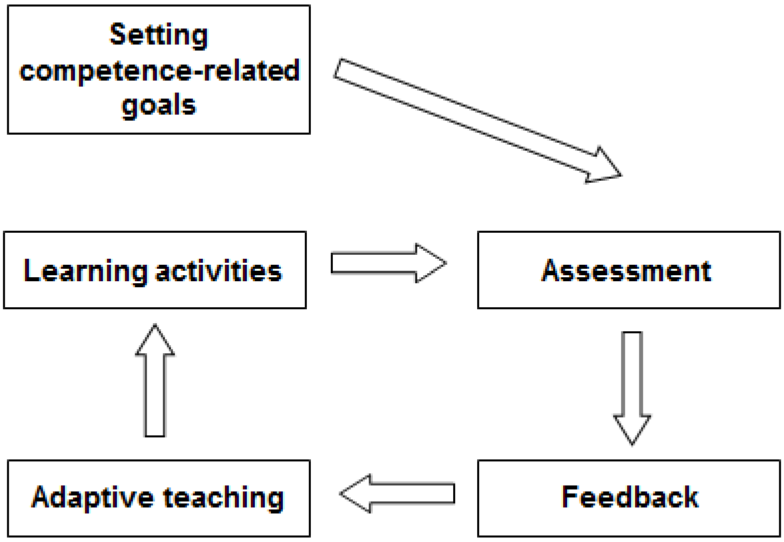

Figure 2: Formative assessment cycle (Klieme & Harks, 2013)

Recent OECD publications stress the importance of formative assessment and the integration of formative and summative assessment, respectively (Looney, 2011; OECD & Centre for Educational Research and Innovation, 2005). They also realize, however, that many logistical barriers to making formative assessment a regular part of the teaching practice exist, rooted in organizational issues such as large classes, extensive curriculum requirements, or the difficulty of meeting diverse and challenging student needs, but also in teachers' often deeply held pedagogical beliefs of assessment as a tool for teacher use and accountability rather than as a method to involve students in a constructivist assessment environment (Wilson & Sloane, 2000).

3. Summary of the literature review

WP 2 has been focusing on synthesizing existing research on formative and summative assessment of IBE related competences in STM. To find relevant publications, an extensive and detailed literature review was conducted. The following sections intend to summarize the findings of this broad and diverse review, guided by the following questions:

- Which aspects of IBE are investigated by empirical studies in STM?
- What formative and summative assessment methods are used in STM with respect to the aspects of IBE?
- How are these methods used?

The review focuses on findings of empirical studies. D 2.4 presents these findings from a comprehensive analysis of existing research on how summative and formative assessment of knowledge and/or competences in STM can be linked to aspects of IBE.

But first, a summary of the procedure of the literature review is given. The procedure considered a search by keywords in relevant databases, an analysis of important journals in STM education (with respect to the assessment of IBE related competences), and a follow-up of main citations. In addition, experts were asked to list relevant publications in the field of IBE as well as in the field of formative and summative assessment. This step aimed at the validation of the search strategies in order to ensure that no relevant publications were missed. Furthermore, all partners from the ASSIST-ME project contributed national reports on the state of the art in their countries (see section 3.6 State of the art in the ASSIST-ME partner countries).

3.1 Procedure of the literature review

This chapter gives a brief summary of the whole procedure which is described in detail in D 2.4. In order to find relevant publications in view of the objectives, three search strategies were realized:

1. Searches in data bases,
2. Searches in relevant journals,
3. Searches in reference lists.

Two data bases were selected for the literature review: 'Web of Science' and 'Education Resources Information Center' (ERIC). The first one has a focus on science journals in general and the second one more on educational journals and materials. Both data bases allow for the systematic and simultaneous search in a collection of most of the important journals within a specific field of interest, e. g. Review of Educational Research, Learning and Instruction, American Educational Research Journal, Journal of Research in Science Teaching, Science Education, International Journal of Technology and Design Education and Educational Studies in Mathematics. The last 15 years, from April 1st 1998 till April 1st 2013, were chosen as the time span.

For the searches in the data bases, certain topics encompassing several related keywords were combined to achieve a high correlation between the content of the literature found and the objectives of the ASSIST-ME project. Table 2 shows the keywords for science education for five out of six topics. By the application of this search with five topics, 331 publications in science, 88 in mathematics and 68 in technology were ultimately found.

Table 2: Keywords and topics for the search in databases in science education

Topics	Keywords for science
Topic 1: Inquiry	Inquiry-based learning OR inquiry OR collaborative learning OR discovery learning OR cooperative learning OR constructivist teaching OR problem-based learning OR argumentation
Topic 2: Subject	science education OR science instruction OR science teaching and learning
Topic 3: School	classroom OR teacher OR student
Topic 4: Objective	assessment OR evaluation OR validation OR achievement OR feedback
Topic 6: method of assessment	discourse OR effective questioning OR assessment conversations OR accountable talk OR quizzes OR self-assessment OR peer-assessment OR portfolio OR learn log OR mind map OR concept map OR rubrics OR science notebook OR multiple-choice OR constructed-response OR open-ended response

Table 3: Relevant journals and their impact factors

Subjects	Journals	Impact factor[1]	
		Last year	Last five years
Science	Journal of Research in Science Teaching	2.55	3.23
	Science Education	2.38	2.71
Technology	Int. Journal of Technology and Design Education	0.34	0.42
	Journal of Technology Education	-	-
	Journal of Technology Studies	-	-
Mathematics	Educational Studies in Mathematics	0.77	-
	Int. Journal of Science and Mathematics Education	0.46	-
	Journal for Research in Mathematics Education	1.55	2.08
Assessment	Applied Measurement in Education	0.58	0.74
	Assessment in Education	-	-
	Educational Assessment	-	-
[1]according to Thomson Reuters, 2013			

In addition to the searches in the data bases, searches in relevant journals (see Table 3) were conducted. The chosen journals were considered as important or even as the most important for each subject or research field with respect to the objectives of ASSIST-ME. Most of them are also included when searching within the two databases.

The articles of all issues published during the last 10 years were scanned by using the homepages of the publishers and the two data bases mentioned. This search strategy was able to improve the quantity and quality of the literature basis. In total, 158 different publications were identified by this search strategy.

To guarantee that important literature with regard to IBE and formative or summative assessment was considered, an additional, more unsystematic search was carried out. Following the pyramid scheme, the reference lists of the literature found were scanned in view of frequently recurring publications that might have a high impact on research on inquiry-based education and formative or summative assessment. For science, there were 32 additional references, for mathematics 10, and for technology and assessment none.

The literature collected by the different search strategies and searches included 701 different publications. Only publications focussing on students in primary, lower secondary, and upper secondary school were taken into account when analysing the reported results. Besides, only publications that present results from empirical studies were regarded. Only 191 of the found publications met these requirements (see Table 4). Even though there was a partial selection before, 510 of all 701 publications were excluded.

Table 4: Final extract for the literature review

Category	S	M	T	Total
Focus students (school)	*148*	*30*	*13*	*191*

The 191 references are publications that meet the objectives of the ASSIST-ME project and formed the final extract for D 2.4. The following sections summarize the most important results of these publications. In interpreting the results, one has to keep in mind that there are only 13 studies in technology and 30 in mathematics, but 148 in science. This made it difficult to determine subject-specific main focuses, especially in technology and mathematics. Furthermore, it was difficult to draw general conclusions. Nevertheless, the findings are regarded as important for the prospective work within the ASSIST-ME project.

3.2 State of the art of IBE

The relevant literature identified by the search in major journals and databases was analysed using a detailed coding scheme. At the core of this coding scheme, specific aspects of inquiry provided the focus when reviewing the articles. The aspects are a comprehensive summary of the diverse definitions of IBE as discussed in D 2.5 (see Table 5).

All of these aspects where found in the reviewed articles, indicating the broad and expansive focus that can be and is applied to IBE in the research literature. However, huge differences were found regarding the extent to which the single aspects are researched, ranging from two (dealing with uncertainty) to 106 publications (argumentation/reasoning) which incorporated a particular aspect of IBE.

Table 5 shows the distribution of reviewed articles across the identified aspects of IBE, in total as well as for each of the three domains. In general, the distribution of researched aspects mainly reflects the distribution in the science domain, as most reviewed articles stem from this subject. Therefore, it is difficult to identify 'well researched' aspects of IBE in all three domains.

In addition, often the majority of the publications include a particular step of IBE only as a facet of a learning environment, while only few studies tried to explicitly assess students' ability on this step. For instance, students' ability to identify research questions was addressed in 44 publications, but in less than one third of these publications students' ability on this step was assessed. Similarly, 26 studies could be identified to incorporate students' dealing with alternative or multiple solutions, but again, incorporating this facet of inquiry was mainly achieved within a learning environment, probably because of the high complexity of the analysis when part of the assessment.

Next to differences between incorporating particular aspects of IBE either as part of a learning environment or as part of an assessment, some aspects of IBE are incorporated quite universal or even as a buzzword only. For instance, students' ability to formulate hypotheses or research conjectures was explicitly addressed in 38 publications. Despite this large number of studies, only few disentangled this aspect of inquiry in detail. Several studies only recorded whether or not the students made assumptions or formulated hypotheses, but did not put further emphasis on analysing the amount or quality of these assumptions. Similarly, the evaluation of results by students is included in many publications as a step of inquiry, but often only as a buzzword or by-product of a more general view on inquiry.

The most researched aspects of IBE are argumentation and debating with peers. All in all, 106 publications included aspects of argumentation, constructing and critiquing arguments or explanations. Among these studies, both the fostering of students' content knowledge by improving their argumentation skills and the fostering of argumentation skills as a worth on its own can be found. Regarding facets of communication processes, 70 studies were identified. Here again, the majority of these studies included this facet of IBE only as part of the learning environment. Interestingly, several studies that included communication as part of the assessment tended to analyse written artefacts.

In contrast to these broadly researched aspects, some blind spots are to be mentioned. Only few studies focus on students' abilities to find structures or patterns. This might be due to the fact that it cannot be clearly separated from e.g. 'searching for generalizations' in mathematics or 'collecting and interpreting data' in science, respectively. However, students' searching for generalizations is included in a learning environment in only five studies, as part of the assessment in only one study. Even less frequently researched is students' dealing with uncertainty. Only two studies were identified including this aspect of inquiry.

Table 5: Distribution of reviewed articles across the identified aspects of inquiry

	Total	M	S	T
scientific inquiry and/or science process skills	55	2	48	5
diagnosing problems and/or identifying questions	44	6	36	2
searching for information	17	1	16	0
considering alternative or multiple solutions	16	1	12	3
creating mental representations	11	5	6	0
constructing and using modeling	17	2	11	4
formulating hypotheses	31	1	30	0
planning investigations	39	2	36	1
constructing prototypes	12	0	4	8
finding structures or patterns	11	4	7	0
researching conjectures	6	1	5	0
collecting and interpreting data	43	1	42	0
evaluating results	38	5	31	2
searching for alternatives and/or modifying design	10	0	7	3
constructing and critiquing arguments or explanations, argumentation, reasoning, and/or using evidence	106	13	91	2
debating with peers and/or communication	70	7	62	1
searching for generalizations	8	4	4	0
dealing with uncertainty	2	1	1	0

Note. The height of the bars is normed to the maximum within each column

Next to these blind spots, several aspects have seldom been researched. Only few studies, for instance, focused on students' search for information, especially as facet of the respective assessment procedures, and they were almost exclusively located in the field of science education. The use of mental representations seems to be a characteristic feature of mathematics and science education. Among these two domains as well as between the adoption of mental representations as part of the learning environment or the assessment, the studies extracted in these reviews are almost evenly distributed. Additionally, students' ability to construct and use models was explicitly addressed in only 17 publications.

Despite the broad definition of inquiry which led the focus of this review, several publications included further aspects. Some of these aspects are domain-specific, for example proof competence as part of inquiry in mathematics education or constructing prototypes in technology education. Representing data by graphs, visualizing data, drawing, and graphing, or using visualizations in general are also partly linked to mathematics, but without doubt these aspects are relevant for the domains of science and technology, too.

In addition, epistemological aspects are addressed in several publications. Epistemic understanding was either regarded as domain-specific, e.g. the nature of science, or as more general, epistemic understanding or the nature of modelling. Interdisciplinary relevance is also enclosed in abilities like divergent thinking and creativity or critical thinking. However, these aspects are not only limited to the domains of STM, but are more closely related to aspects of general cognitive abilities.

Beyond these cognitive abilities, also affective aspects are addressed in different publications. As theses aspects were not in focus of the literature review and consequently relevant terms were not included in the database search, these affective aspects were found only to a small extent. Enjoyment, interest, value, self-efficacy, motivation, and confidence, but also attitudes towards science or mathematics are analysed in relation to different aspects of inquiry.

Table 6 summarizes those aspects of IBE that are incorporated in at least 10% of the research papers within each of the three domains (to account for the large differences in the amount of reviewed articles between STM).

In mathematics education, the most researched aspects are diagnosing problems, creating mental representations, evaluating results, argumentation, and debating with peers. These aspects reflect a picture of inquiry in mathematics education[1] that emphasizes the processing of real-world problems with mathematical tools and models. This includes the transfer of the real-world problem to a mathematical question (e.g., by diagnosing problems and creating mental representations) and the transfer backwards from the mathematical world to the real world (e.g., by evaluating results, i.e. by con-

[1] The interpretation reflects the findings of the literature review, i.e. it is founded on empirical, descriptive data. No normative emphasis regarding the relevance of particular aspects of IBE within each domain, especially those that are researched less frequently, is intended by the authors.

necting the numerical or functional expression gained from mathematical operations to the initial situation). Accordingly, inquiry in mathematics education is closely linked to steps of problem-solving as already emphasized by Polya (1945).

Table 6: Aspects of IBE that are incorporated in at least 10% of the research papers within each of the three domains

Science education	Technology education	Mathematics education
diagnosing problems and/or identifying questions		diagnosing problems and/or identifying questions
	considering alternative or multiple solutions	
		creating mental representations
	constructing and using models	
formulating hypotheses		
planning investigations		
	constructing prototypes	
collecting and interpreting data		
evaluating results		evaluating results
	searching for alternatives and/or modifying design	
argumentation, reasoning, and/or using evidence		argumentation, reasoning, and/or using evidence
debating with peers and/or communication		debating with peers and/or communication

In science education, the majority of reviewed articles focus on similar aspects (identifying questions, evaluating results, argumentation, and debating with peers), but additionally incorporate the aspects of formulating hypotheses, planning investigations, and collecting and interpreting data. The sequence of question, hypothesis, investigation, data collection, and results resemble a common pattern of scientific experimentation. This sequence is often followed, accompanied or even primed by argumentation and the debate with peers.

In technology, the most frequent aspects of IBE in the reviewed articles seem to reflect a strong product-orientation (e.g., constructing and using models, or constructing prototypes) and facets of trade-off and balancing between different opportunities (e.g., considering alternative or multiple solutions, or searching for alternatives and/or modifying design). Regarding the similarities between the most researched aspects in the three domains, the commonalities between mathematics and science seem to be greater, while technology education puts a different emphasis in its interpretation of IBE.

3.3 State of the art of formative and summative assessment

One has to note that the focus of the search strategy of the literature review was on IBE in STM in combination with assessment methods. Therefore, most of the studies using or investigating assessment methods have to be seen against the background of IBE and related aspects and competences. Of course, there are many more studies describing or exploring assessment methods in general or in STM, but they were intentionally not subject of the literature review.

For the analysis of the assessment practices in IBE in STM, the frequency of the assessment types used was compared between science, technology and mathematics. In three quarters of all studies, methods of summative assessment were employed. Methods of formative assessment were not very common among the empirical studies found, especially in science education. Peer- and self-assessment played a subordinate role. In combination with IBE, neither was explored very often. In contrast, rubrics were a common instrument used for the evaluation and analysis of varying assessment situations.

In view of the objectives of the ASSIST-ME project, it is important to know which assessment methods are frequently employed in the studies and which assessment methods are less common. Furthermore, the purpose of the assessment methods is of importance. In the following paragraphs, these aspects are summarized for each subject.

3.3.1 Assessment practices in science

Multiple-choice and constructed-response or open-ended items used as a summative assessment tool dominate the assessment methods in research on IBE in science education. The reasons are obvious as these items have many advantages. In particular, the analysis of multiple-choice items is more objective and the results are easier to compare and to interpret than other more complex assessment methods. Even though the items have advantages in view of summative assessment, they are less frequently used for formative assessment. One example is provided by Hickey and Zuiker (2012) who used open-ended items to support feedback conversations.

Among the found publications, quizzes were only used by one research group (e. g. Hickey, Taasoobshirazi, & Cross, 2012). Ultimately, the quizzes developed by Hickey et al. (2012) were a combination of multiple-choice and open-ended items. First, the students had to give a short answer and then an explanation to support that answer.

To assess students' understanding of key concepts, concept maps instead of items are often used for a summative assessment (e. g. Brandstädter, Harms, & Großschedl, 2012). On the other hand, concept maps can be used for formative assessment as well. In this case, the focus lies on checking students' progress in understanding key concepts at several times during a treatment (e. g. Furtak et al., 2008). One possibility to organise the analysis of concept maps are rubrics (e. g. Nantawanit, Panijpan, & Ruenwongsa, 2012). Another mapping technique is mind mapping. It is a tool that can be used to ascertain students' developing ideas about scientific concepts (Goodnough & Long, 2006). Similar to concept mapping, the technique makes the exploration of prior knowledge possible, as well as an assessment of students' overall performance from the viewpoint of specific learning outcomes.

Hands-on activities like experiments are often used for performance assessments in a summative manner. They are supposed to be an alternative to more traditional paper and pencil assessment methods like multiple-choice items (Shavelson, Baxter, & Pine, 1991). For example, the engagement in inquiry-type experiments in the chemistry laboratory improves students' ability to ask high-level questions, to hypothesize, and to suggest questions for further experimental investigations (Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005). In this case, the experiments were a method to provoke a more realistic assessment situation. Among the publications, there was one study which fully met the objectives of the ASSIST-ME project (Pine et al., 2006). By conducting a performance assessment, the inquiry skills 'planning an inquiry', 'observation', 'data collection', 'graphical and pictorial representation', 'inference' and 'explanation based on evidence' were measured.

However, in comparison to multiple-choice or constructed response items, performance assessments require more complex scoring or evaluation systems. Therefore, notebooks are an alternative mainly used in formative assessment. They are supposed to monitor and facilitate students' understanding of complex scientific concepts and especially inquiry processes. To achieve this, the method includes the collection of student writing before, during, and after hands-on investigations (Aschbacher & Alonzo, 2006). Thus, notebooks can obtain information about students' understanding at any point in the curriculum without needing additional time. Moreover, Baxter, Shavelson, Goldman, and Pine (1992) were able to confirm that notebooks are also a valid tool for a summative assessment of hands-on activities. However, in view of performance assessment, notebooks are a reliable tool that can be used for formative teacher feedback (Ruiz-Primo, Li, Ayala, & Shavelson, 2004). Another method for a summative assessment of investigations or inquiry processes, are portfolios which summarize the whole process (e. g. Dori, 2003). Portfolios are normally compiled individually to measure knowledge or competence growth over a certain period of time.

Usually, conversations or discussions are a method to enhance students' argumentation, reasoning or communication skills. Mainly, discussions take place in small groups. An approach which supports collective discourse is the feedback conversation (Hickey et al., 2012; Hickey & Zuiker, 2012). It suggests that the most valuable function of feedback is fostering participation in discourse. Apart from a formative character, one can use discussions with a more summative character with regard to the assessment. For example, Mason (2001) used students' small group discussions to address aspects of IBE, e. g. formulating and comparing hypotheses. Instead of organizing discussions around phenomena, students can also discus socio-scientific issues (e. g. Nielsen, 2012; Osborne, Erduran, & Simon, 2004) in order to balance different positions on a certain topic. This kind of issues calls for a discussion about what to do and not merely about what is true. Closely related to discourses, assessment conversations or accountable talks can also be employed as assessment methods (Ruiz-Primo & Furtak, 2006).

Communication processes are often observed, for example, to assess students' argumentation within discussions or classroom interaction (e. g. Abi-El-Mona & Abd-El-Khalick, 2006). Shemwell and Furtak (2010) investigated the quality of argumentation in classroom discussion by analysing the support of argumentation by evidence. In another study, McNeill (2009) analysed the instructional practices teachers use to introduce scientific explanations by videotaping classroom interaction. Moreover, observations provide records of the order in which students carried out certain activities in learning environments and the time they spent on these activities (e. g. Kubasko, Jones, Tretter, & Andre, 2008). For some reasons, it is necessary to combine both purposes (e. g. Harskamp, Ding, & Suhre, 2008). Another purpose is the observation of students' performance in a certain task (Sampson, Grooms, & Walker, 2011).

Methods that can be used for observations are for example field notes or video and audio tapes. The application of video and audio tapes aims more at the observation and analysis of learning and teaching processes than at the assessment of learning or teaching outcomes (Valanides & Angeli, 2008), even though they are generally used for summative assessment. Moreover, they are used as a further tool in addition to other research methods or in explicit combination with other tools (e. g. Vellom & Anderson, 1999).

In cases in which only audio tapes were used, the focus was on the talk especially on the amount of on/off task talk and the categorization of task talk (Cavagnetto, Hand, & Norton-Meier, 2010). Chin and Teou (2009) audiotaped conversation from one group of students to provide a record of students' thinking in a form that was accessible to the teacher for monitoring and feedback purposes. This is an example of a formative use of audio tapes. Students' assertions and questions had formative potential as they encouraged discourse by drawing upon each other's ideas.

In addition, field notes are a method which combines both observations and video or audio tapes. For instance, they provide general descriptions of the most salient instructional events during an observed session (e. g. Abi-El-Mona & Abd-El-Khalick, 2006) or provide information about events that occur outside the range of a video camera (e. g. Ryu & Sandoval, 2012). Furthermore, field notes can be taken as events unfold, and recorded with time indices for later matching with video segments (e. g. Vellom & Anderson, 1999).

Similar to any kind of observation, the objectives of interviews are also manifold and, similar to field notes, they are an additional tool that is usually combined with other methods such as observation, video tapes (e. g. Berland, 2011) or audio tapes (Dawson & Venville, 2009). Interviews are an assessment and research method that is usually qualitatively analysed. For example, after responding to a questionnaire, students were asked to explain their answers in order to gather information about existing misconceptions (White & Frederiksen, 1998). Furthermore, pre- and post-interviews provide another possibility for evaluating the intervention part of a case study (Berland, 2011). Ash (2008) gives an example of how interviews can be used as a kind of formative assessment in order to measure students' competence in solving biological dilemmas. An interviewer provided biological dilemmas as thought experiments, described the context, and then asked questions. The formative character was introduced by further questions or hints: After the student had answered, the interviewer provided a hint if the student was on the wrong track or a challenge if the student gave an appropriate answer. The hint determined what a student might achieve with appropriate help, while the challenge helped to determine whether the understanding was robust.

3.3.2 Assessment practices in technology

In total, empirical studies on IBE in technology education in combination with assessment methods are rare. Obviously, in contrast to science education, this research field is not particularly dominant. One reason might be that technology is not a common subject in European schools (see D 2.3, National reports of partner countries reviewing research on formative and summative assessment in their countries) or in American schools. One effect of the small number of publications is that there is a limited variety of assessment methods used within the reported studies. The prevailing number of methods was not used, e. g. concept map, mind map, learn log, note-book, effective questioning, video tapes, or artefacts.

With regard to summative assessment, the most important methods are, similar to science education, constructed-response or open-ended items and multiple-choice items. In most cases, they were used for the assessment of knowledge, achievement or understanding.

When looking at formative assessment, the most important methods are portfolios and interviews. Obviously, the advantage of portfolios is their ability to reconstruct a process when solving a problem or designing a prototype (Barak & Doppelt, 2000; Doppelt, 2003; Hong, Yu, & Chen, 2011). Interviews are used to probe students' understandings of materials and stability (Davis, Ginns, & McRobbie, 2002).

Another realized assessment method is performance assessment in order to compare students' competence in hands-on and virtual construction of a prototype (Klahr, Triona, & Williams, 2007).

3.3.3 Assessment practices in mathematics

As well as in technology education, the number of publications in mathematics is quite small. One reason might be that IBE is not a common approach in mathematics education. Its province is the approach of problem-solving.

In mathematics, the emphasis is on constructed-response or open-ended items – especially for summative assessment. The purpose of the items was often the evaluation of an intervention by a pre-post-design. The items ascertained students' reasoning or problem-solving skills and their mathematical knowledge. The use of constructed-response or open-ended items is not surprising, as in mathematics education, students usually have to calculate and write down the calculation or prove and explain a given problem. Among the studies, Heinze, Cheng, Ufer, Lin, and Reiss (2008) gave examples of test items which measure students' proof competence. Others assessed problem-solving skills (e. g. Schukajlow et al., 2012) which also involve proof competence. In contrast to science and technology education, multiple-choice items are less common in mathematics education. It is assumed that they would simplify the tests by providing different answer options. Therefore, they are not regarded as suitable for the assessment of problem-solving skills.

Another emphasis is on the observation of lessons or learning situations by observations, field notes, video tapes and audio tapes. The application of these methods was often not described in detail. As these methods were used in a more qualitative way, the focus of the respective publications was on the description of the observed learning or teaching processes (e. g. Boaler, 1998). Other studies focused on the analysis of discourse, assessment conversations or accountable talk in connection with collaborative learning (e. g. Pijls, Dekker, & van Hout-Wolters, 2007).

The GPAR (Goals, Plan, Action, Reflection) reflection sheets published by Brookhart, Andolina, Zuza, and Furman (2004) are different from all other assessment methods because of the combination of reflection and feedback. They ask students to write responses to given questions in order to reflect their own learning process and to give feedback immediately. Therefore, this method can be useful in view of formative assessment.

The found publications did not focus on methods that are mainly used formatively, e. g. concept map, mind map, learn log, notebook, and quizzes. Obviously, there is a need for more research on formative assessment in connection with IBE in mathematics learning.

3.4 Connections between IBE and assessment methods

The detailed description of both the different aspects of IBE and the different methods used for the assessment of IBE provides a rich picture of the different possibilities and diverse focal points in both areas. However, it is difficult to identify connections between single aspects of IBE and predominantly used methods to assess this specific aspect. Nevertheless, the question of common and less common connections between aspects of IBE and assessment methods is important when planning new interventions as it is the case in the ASSIST-ME project. On the one hand, common connections (in the sense of 'used often') indicate that there is a lot of research literature and thus a lot of examples for assessing aspects of IBE with specific assessment methods. On the other hand, missing connections between particular aspects of IBE and specific assessment tools might indicate a mismatch between both, but might also indicate blind spots that might be worth further investigation.

In order to identify these connections between the focus on different aspects of IBE and specific methods of assessment, the results of the literature review have been used. Accordingly, the following pictures of networks (see Figure 3 to Figure 5) created with 'R' (R Core Team, 2013) and the 'igraph package' (Csardi & Nepusz, 2006) display all connections that were found in the respective studies of the review. Each combination of one aspect of IBE and one assessment method (cf. Table 9 in D 2.4) is represented by one connection in the networks. When more than one aspect of IBE was included in a single paper or when one aspect of IBE was analysed using several methods of assessments, every combination was included as one relation. Hence, there are more connections in the networks than publications in the literature review.

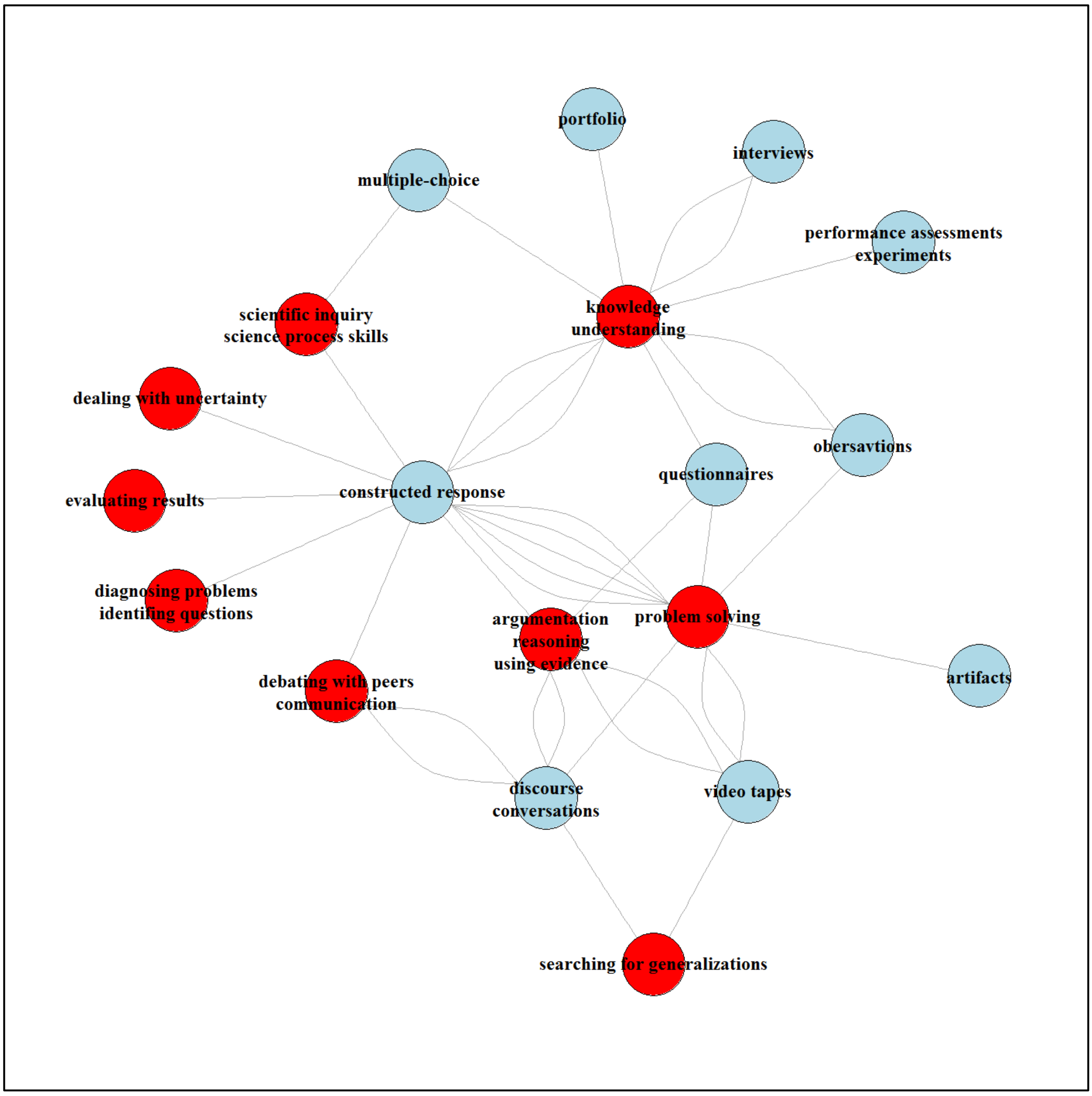

Figure 3: Interrelations between aspects of IBE (red) and assessment methods (blue) in mathematics education (based on the reviewed articles)

The data from the literature review is purely descriptive. Accordingly, the networks are only able to display the interrelations between aspects of IBE and assessment methods based on the sample of articles that were included in the review. The more connections are drawn between aspects of IBE and methods of assessment, the more often this combination was used to assess particular aspects of IBE with a specific method of assessment in the articles of the literature review. This does not mean, however, that missing connections or combinations which are rarely represented in the networks could not also be reasonable combinations. This constraint has especially to be taken into account when regarding the networks based on the reviewed articles in mathematics and technology education, as only a small number of articles had been included in the review.

With regard to the interrelations between aspects of IBE and assessment methods in mathematics education (see Figure 3) it becomes obvious that the use of constructed responses was often and widely used in the reviewed articles to assess diverse aspects of IBE. Additionally, the more general aspects ‘problem solving’ and ‘knowledge and understanding’ were included in several studies, often as the dependent variables to which the researched aspects of IBE were related. Focusing on specific aspects of IBE, the aspect of ‘argumentation, reasoning and using evidence’ is investigated quite often and assessed with the use of different methods.

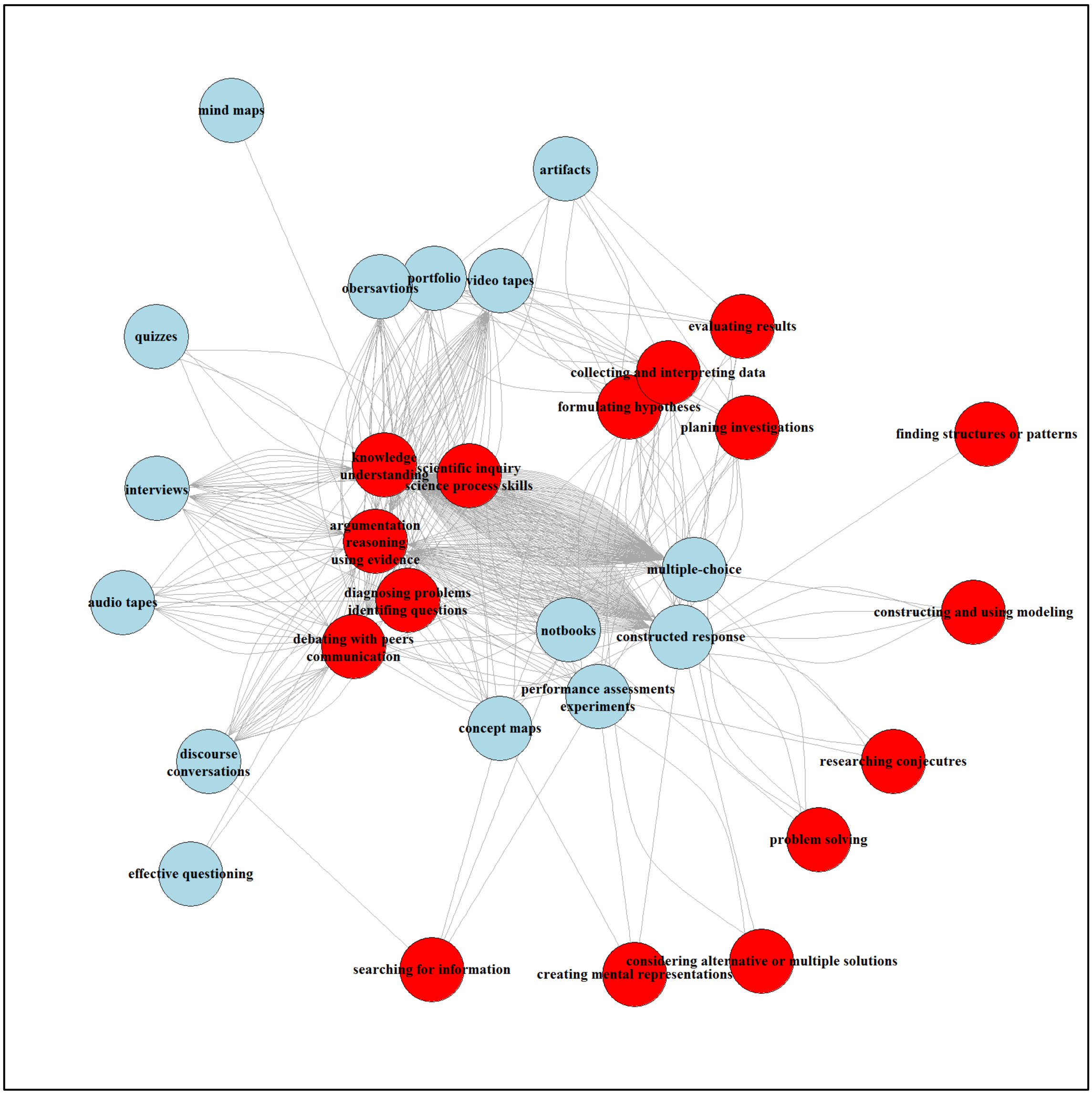

Figure 4: Interrelations between aspects of IBE (red) and assessment methods (blue) in science education (based on the reviewed articles)

With the small data set of reviewed articles in mathematics education in mind, it is difficult to generalize the results beyond these reviewed articles. Generally, a broad array of IBE aspects is researched with different methods, but it is not possible to identify specific patterns. The found interrelations as displayed in Figure 3 are expectable, for instance that aspects like ‘debating with peers’ are mainly assessed by open formats like ‘discourse conversations’.

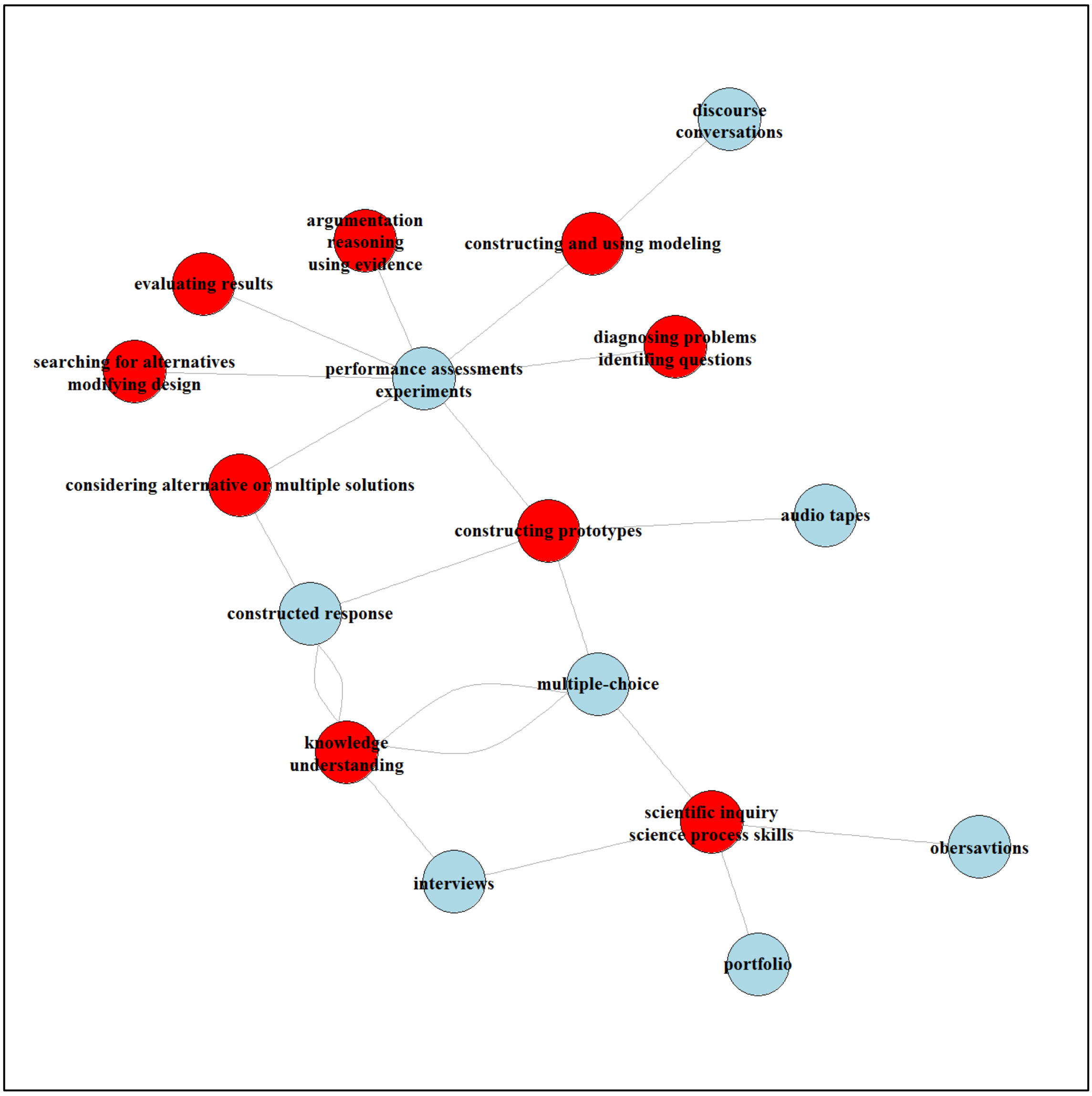

Figure 5: Interrelations between aspects of IBE (red) and assessment methods (blue) in technology education (based on the reviewed articles)

A different picture is received in science education. Here, the sample set of reviewed articles is much larger, thus revealing common and less common combinations of IBE aspects and methods of assessment (see Figure 4). First, the graph resembles the impression already gained in the descriptive report on aspects of IBE and assessment methods (D 2.4). The aspect of ‘argumentation, reasoning, and using evidence’ is researched most often, followed by ‘knowledge and understanding’ and ‘scientific inquiry’ as general aspects and ‘diagnosing problems’ and ‘debating with peers’ as more specific aspects of IBE. These aspects are assessed by a broad array of methods, in particular with the aid of multiple choice and constructed response instruments, but also by using observations, portfolios and video tapes. The aspects of ‘planning investigations’, ‘formulating hypotheses’, ‘collecting and interpreting data’ and ‘evaluating results’ that are central aspects of a classical inquiry process are investigated less often. Here again, different methods of assessment are used, including written formats and observational methods. All other aspects of IBE that were included in the literature review are assessed less often and then only with one or two different methods.

With regard to Figure 4, it is seems also to be interesting to look for 'missing links'. For instance, 'mind maps' has received increasing attention both in educational research as well as in teaching practice. The creation of maps is intended to represent knowledge structures and to enable students to derive connections between content aspects. However, mind maps are only used once in the reviewed articles and in this case to assess students' knowledge. No combination of mind maps with aspects of IBE like 'creating mental representations', 'constructing and using models' or 'researching conjectures' have been found.

In the field of technology education, again the small data set of reviewed articles constrains the identification of common and less common interrelations between aspects of IBE and methods of assessment (see Figure 5). Here, the aspect 'performance assessments and experiments' is the only knot that is bound to a noteworthy number of different aspects. In general, only one or two connections occur in the graph, certainly due to the small number of reviewed articles in this area.

3.5 State of the art of ICT

As part of the WP 2 literature review, Pearson Education International analysed the use of e-assessment in formative and summative assessment of STM subjects with a focus on inquiry-based and competence-based learning (D 2.4 – part II). The aim of this review was to inform the development of digital assessments that are relevant to the aims of ASSIST-ME. It looked at different aspects of e-assessment, IBE and summative and formative assessment, respectively, and is organised around four objectives:

1. identifying theories and models which are relevant to the development of digital assessments,
2. identifying strategies used in the evaluation of the models which could inform good practice,
3. identifying existing relevant digital assessments,
4. identifying implications for the development of the digital assessments relevant to the aims of ASSIST-ME.

In general, the majority of research in e-assessment was found to be located at the higher (HE) and further education (FE) level. However, the results and the general principles nevertheless provided information that could be used to inform and support the further work in the ASSIST-ME project. Only few publications were found that explicitly dealt with the use of e-assessment in IBE. Examples cited are: 'Operation ARA (formerly ARIES!)' in which learners are presented with inaccurate scientific information and must ask questions to ascertain the truth (Halpern et al., 2012); 'SimScientist' which is an interactive simulation-based science assessment (Pellegrino & Quellmalz, 2010); the NAEP 2009 framework for science and the NAEP 2014 framework for technology and engineering literacy (Wylie & Dolan, 2013); the draft PISA 2015 framework for scientific literacy (OECD, 2013); and 'Scholar', a secondary science and mathematics environment, including simulations, that utilises animated graphs which could be used as stimuli for IBE (Scholar, no date).

3.5.1 Theories and models in e-assessment

With respect to the first objective, the report identified a variety of theories and models for e-assessment. Compared to traditional paper-and-pencil tests, e-assessments have several advantages:

- E-assessments facilitate the link between teaching, learning and assessment. The identified models (e.g. gaming-style models) have successfully blended teaching and short, repetitive formative assessments and demonstrated how these tasks can be combined to give a final, more summative assessment.
- In e-assessments a variety of stimuli can be used including digital learning objects, worked exemplars, simulations or real-life scenarios. Students can interact with these stimuli what consequently increases student motivation.
- Many successful models of e-assessment include feedback. The feedback may take a variety of forms from simple to complex. One of the major advantages of digital assessments is seen in the immediacy of the feedback which allows students to react on feedback 'there and then' and thus is beneficial for student motivation.
- E-assessments allow for adaptivity (e.g. by computer adaptive testing – CAT) which enables a more interactive approach for learners. Learning and assessment become personalized. In combination with feedback this may lead to group work and peer-assessment thus increasing learner autonomy and motivation.
- E-assessments allow for the collection of large amounts of data from simple scores to progress information, insights how learners approach a task and information about e.g. misconceptions. This can enhance the teacher's understanding of her/his students' learning and thus act as a strong contributor to formative assessment.
- E-assessments allow for the re-submission of answers. The immediacy of feedback and the possibility to re-submit makes learning more relevant for the students. Moreover, it allows them to move directly on to the next stage which has advantages in formative assessment and with self, peer and diagnostic assessment. Learners can also submit a confidence rating with their answer which adds further meaning to the level of information gathered.

3.5.2 Evaluation of e-assessment models

The review only found few examples of how e-assessments had been evaluated (objective 2). A model for development of e-assessments did not seem to follow the usual pattern of design – trial – evaluation and so few cases of evaluation techniques were identified. In the 'Operation ARIES!'-project, learning gains were measured using short answer responses and compared across different types of e-learning and the immediacy of feedback. The evaluation of the ALTA system (Adaptive Learning Teaching Assessment for mathematics KS1–3; ALTA, 2009) distinguished between practical concerns (like e.g. the ease of use and teacher training) and educational issues (like e.g. student engagement and the development of skills).

3.5.3 Exemplars of e-assessment

With respect to objective 3, the review identified several well-established e-assessment programmes in the field of STM. These programmes show how e-assessment can be integrated into the learning process and successful in terms of motivating learners and assessing competencies that are more difficult to assess using paper-and-pencil tests. Moreover, e-assessments offer the possibility of including elements difficult to replicate in a paper test, for example, simulations, videos and scenarios that lend themselves more readily to IBE and competency-based skills.

The mathematics exemplars presented in the review included the use of questions, simulations, enrichment activities, activities designed to assess misconceptions and online games. They could be used in summative assessment, formative assessment, self-assessment and diagnostic assessment and covered a wide range of age groups from 5 year-olds to higher education. The science exemplars focused on summative and formative assessments and were aimed at ages 11-19. They often involved inquiry tasks and processes of critical thinking and incorporated the use of feedback. Other characteristics are that they were interactive, utilised e-coaching and intelligent tutoring, offered possibilities to re-submit answers, and employed videos, case studies and simulations. The technology examples found were used summatively as they were all high stakes qualifications (aiming at 14-19 year olds). However, some aspects could also be used for formative assessment purposes. The exemplars required students to solve problems in a virtual world, use interactive simulations and intelligent tutoring/online tutor support. They allowed students to progress at their own rate and included the use of a timeline of student activities and an e-portfolio.

3.6 State of the art in the ASSIST-ME partner countries

To assess the state of the art of formative and summative assessment in STM in the ASSIST-ME partner countries, researchers from each country were asked to collect national research results based on ten questions provided by WP 2 as D 2.1. The complete country reports are summarized in D 2.3. In the following, some of the major findings from the national reports will be presented.

In some respect, the most important outcome of these reviews for the further work in the ASSIST-ME project is that almost all countries stated that there has been little to no research on formative assessment – or assessment in general – in their countries. One reason for this was e.g. seen by Cyprus in the fact that evidence-based research is not prioritized in policy decisions. Despite the general lack of research results, however, most countries tried to answer the questions provided in the guidelines by combining the results that were available with their own hypotheses and information about the educational systems relevant in the context of the questions.

All countries report that summative assessment is predominant with the character of the assessments differing from nationwide comparable to school or even teacher-based tests in relation to e.g. the centralization of the educational system or the autonomy of schools. Especially when high stakes are connected to these summative assessments they heavily influence teaching practice and the use of formative assessment, respectively. Moreover, existing large-scale national assessments e.g. in Germany or Switzerland aim at providing information at the educational system and not at the individual level.

Structured formative assessment seems not to exist in the ASSIST-ME partner countries. Even if mandatory guidelines at the policy level exist that explicitly mention formative assessment (like e.g. in Switzerland as frequent, short, straight-forward actions to support student's learning and motivation), there are no systematic surveys of the actual formative assessment practice in schools. There are small- to medium-scale projects or studies in some countries (e.g. Germany, Denmark or Switzerland) which are investigating certain aspects and effects of formative assessment but they do not reflect everyday teaching practice. However, several countries (e.g. Finland, Denmark and Switzerland) report evidence that teachers use informal formative assessment in their regular teaching, especially through questioning or in classroom discussions. Results from France and Switzerland show a high acceptance of formative assessment by teachers, students and parents. In England, some teachers use pre-emptive formative assessment which means that they plan activities and ask questions that they know from experience may cause problems in their students' learning. Such an approach could be seen as scaffolding and supportive of learning. However, it sometimes provides a shortcut which enables learners to complete a difficult task but without providing the opportunity to fully sort out the initial problem in the learning. Also in England, a national pilot called 'Assessing Pupil Progress' (Office for Standards in Education, Children's Services and Skills (Ofsted), 2011) was launched in 2010 which provides a criterion-referenced approach to learning. Although in essence, it is a help to make summative decisions, some teachers manage to use it more formatively.

Countries regarded different factors as impeding the uptake of formative assessment. These factors are mostly in line with the results found in the international literature. A serious impediment is seen in teachers' beliefs about assessment as an instrument for generating grades and ranking students (e.g. in Switzerland). In other countries, e.g. in Cyprus, research shows that although teachers seem to appreciate assessment as an integral component of teaching and a powerful means of enhancing the quality of teaching and learning, they nevertheless exhibit an inclination towards traditional assessment approaches that yield overall scores. Other aspects mentioned e.g. in a study from Switzerland are a lack of time and a lack of teacher competence e.g. to differentiate between different levels of proficiency within a class. Moreover, teachers often seem to have reservations towards formative assessment because they consider it to be laborious and difficult to implement (e.g. in Finland). Studies from England could show that implementing formative assessment is a massive undertaking for schools. Although 'assessment for learning' had been adopted as a National Strategy for whole school improvement which undoubtedly raised its profile, its implementation was hindered by competing priorities at schools, especially the summative agendas. Another observation mentioned in the English context is that external assessment systems seem to undermine not only the assessment skills of teachers but also their confidence in their ability to make sound assessments of their students.

With respect to support that teachers need in order to implement formative assessment into their daily teaching practice, almost all countries agree on a general need for pre- and in-service teacher training that recognises, develops and values the professionalism of teachers. This training might address different aspects. Among the aspects named by the countries are:

- Increasing teachers' 'assessment literacy' (e.g. diagnostic competence, peer- and self-assessment, personal vs. normative frameworks, etc.),
- Changing teachers' beliefs about assessment (see above),
- Increasing teachers' pedagogical knowledge of student learning,
- Developing dialogic classrooms,
- Creating classroom environments that support formative assessment (e.g. by fostering student autonomy and learning readiness),
- Providing teachers with a professional language to discuss the changes they implemented and their impact (e.g. by collaboration and reflective portfolios).

In this context, the importance of a strong relation between educational research and assessment practice is stressed by e.g. Denmark and the Czech Republic. Teaching, learning and assessment should be closely interlinked in the planning and implementation of any teaching programme. An urgent need for concrete assessment tools is expressed in a study from Switzerland. They found that ready-made maths units including rubrics for assessment encouraged teachers to assess complex (and therefore often neglected) competences (Jundt & Wälti, 2011; Smit & Birri, 2012). In Finland, a possible way to support teachers could be to involve textbook writers in the process because of the central role textbooks play in Finnish teaching and learning. From a study on school effectiveness, eventually, Cyprus concludes that mechanisms for internal evaluation need to be established and activities implemented that aim at improving teaching practice and the corresponding learning outcomes.

No studies investigated whether the assessment methods influence the uptake of IBE in the respective countries. Switzerland, Finland, and the Czech Republic explicitly state that IBE is not used frequently, is uncommon or is not a part of the regular instruction, respectively. Different hypotheses exist, however, concerning this topic. For England, it is stated that the emphasis on coursework at GCSE has greatly hampered IBE development since much of the coursework done in schools in STM is ritualistic and narrow in focus. Another major issue is that IBE is often not assessed in examinations (e.g. in Denmark and Finland) and is thus perceived as auxiliary to core teaching. However, there has been some research in Denmark on how summative and formative assessment could be used to promote learning in IBE (the 'assessment dialogue'; Christensen, 2004). A study in Germany (Winter, 2007) points out that a dilemma between alternative assessment methods that aim at the contemplation of learning (like e.g. learning diaries) and student evaluation exists. Students might not openly express their ideas, opinions, and problems if they know they will be evaluated. On the other hand, students might be demotivated if they put much effort into e.g. a portfolio and this work does not contribute to the grades at all. In Switzerland, possibilities for IBE are assumed to exist as no high stakes are connected to assessments. It is hypothesized that with more support for teachers in formative and summative assessment, IBE could gain significance. In a similar way, it is assumed in the Czech Republic that formative assessment facilitating the steps during an IBE activity could improve the adoption of new procedures based on IBE learning. In Germany, the implementation of educational standards (which include IBE competences) required the development of competence models – and thus assessment items – for IBE for monitoring purposes.

With respect to the existence of instruments for formative assessments, countries differ significantly. Whereas in Finland and the Czech Republic no instruments exist at all, e.g. in Denmark the Ministry of Children and Education has published guidelines and instructions for a total of 25 different tools or methods for assessment of which most can be used directly for formative assessment (Ministry of Children and Education, no date). However, very little research-based knowledge exists on how these instruments are used. Another instrument developed in Denmark is called 'The Teacher-Student Assessment Dialogue on Subject Matter' that attempts to integrate assessment and instruction (Christensen, 2004). It was shown to be conducive to student learning and to amplify the learning processes that are prompted by regular teaching. Instruments from England include subject specific booklets to promote formative assessment as well as commercially available multiple-choice questions that can be used with electronic whiteboards or all web-enabled devices (e. g. Socrative.com, 2013). A research group at York is currently developing diagnostic questions in science that may be of use formatively. In Switzerland, formative assessment is systematically gaining importance and being supported by regulations. Examples for formative assessment formats are rubrics, portfolios in mathematics, and textbooks fostering IBE that include assessments. However, they do not reflect the daily practice in school. Similar to Denmark, instruments for formative assessment exist in Germany but there is only little research about their use. Recently, however, several studies investigated the use and effect of feedback in mathematics instruction (Klieme et al., 2010; Rakoczy, Harks, Klieme, Blum, & Hochweber, 2013). A categorization system for feedback was also proposed in France (Georges & Pansu, 2011). Other French studies investigated the use of a computer-based tool for diagnostic assessment in mathematics (Grugeon, Pilet, Chenevotot, & Delozanne, 2012) and of a so-called 'personal contract of success' which is supposed to enable students to identify their learning difficulties and show them ways to improvement (Talbot, 2009). With respect to IBE, several small to medium scale studies in Germany focused on the (summative) assessment of experimentation competence in science education (e. g. Björkman, Labetzki, & Tiemann, 2010; Hammann, Phan, & Bayrhuber, 2007). In contrast, a Danish formative assessment instrument aimed at supporting students in performing inquiry processes in physics (Dolin, 2002). It was shown to increase the motivation, especially of girls, dramatically. In summary, one may say that not in all but in most of the ASSIST-ME partner countries approaches to formative assessment exist that might bear potential for the development of assessment tools within the ASSIST-ME project. Research-based evidence of their use and effectiveness, however, is often missing.

4. Final conclusions and recommendations

The aim of the work of WP 2 in the ASSIST-ME project was to provide a research-based foundation for the future work in the project. Based on this work – and the previous reports from WP 2 – the following conclusions are drawn and recommendations are given that are supposed to guide the development and implementation of assessment methods by WP 4 and WP 5.

4.1 Conclusions

The goal of WP 2 was to investigate the state-of-the-art of formative and summative assessment of inquiry-based education in science, technology and mathematics, both in the international literature and also in the specific national contexts of the ASSIST-ME partner countries. Conclusions from this work can thus be drawn at different levels that are related either to IBE, to assessment or to the link between IBE and assessment, respectively.

Inquiry-based education. Inquiry as an approach to teaching and learning is strongly related to science education. In mathematics and technology education, the term 'inquiry' is used much more seldom. Nevertheless, approaches to teaching and learning that share common goals and characteristics with inquiry-based science education exist but come under different names like e.g. 'engineering design' in technology or 'problem solving' in mathematics education. Even in science, however, no precise, consistent and generally agreed upon definition of IBE exists (although, especially looking at former and on-going EU-projects within the field of IBE, a clear preference for the definition by Linn et al. (2004) can be observed). D 2.5 aimed at providing a definition of IBE suitable for the ASSIST-ME project. The results showed that, naturally, subject-specific differences in the definitions of IBE in science, technology and mathematics exist. Nevertheless, the report also identified aspects of IBE that are relevant in all three subjects. Looking at these aspects in more detail, it became obvious that they are not solemnly related to competences but that even well-known and used definitions of IBE often not distinguish between competences, skills, activities and attitudes. Moreover, aspects that are called the same might have slightly different meanings in the different subjects. One important result from the country reports on national research on assessment of IBE (D 2.3) was that the degree to which IBE is a known and used approach to teaching and learning in their respective countries varies significantly. There are ASSIST-ME partner countries in which IBE is practically not practiced at all in daily instruction which may have implications for later phases of the project (e.g. the implementation phase). Summarizing the conclusion with respect to IBE, a need is seen for the project to clearly and precisely define what should be understood under IBE in the context of ASSIST-ME. Another variable that might be important in this context is the degree of openness in the inquiry instruction that can vary from very structured, teacher-guided to completely open formats.

Formative and summative assessment. As with the definition of IBE, there is also some variation with respect to definitions of formative assessment in the literature. Moreover, it has to be kept in mind that formative assessment approaches might come under different names in different countries (e.g. in Germany teachers prefer speaking of ‘individual support’ instead of ‘formative assessment’). This might become important in later stages of the project in order to help teachers identify with the project and its objectives. Although theoretical papers on formative assessment exist, empirical studies mainly focus on summative assessment even if they employ assessment methods that have the potential to be used formatively. These empirical studies are heavily dominated by the use of multiple-choice and constructed or open-response items, respectively. Only few papers have been found that focus e.g. on effects of specific formative assessment instruments or the correlation between teacher characteristics and the use of formative assessment – here a definite need for further research could be observed. The results from the literature review do not straightforward allow for the exact definition of assessment methods or approaches. Multiple-choice items, for instance, can be constructed and used for completely different purposes like the assessment of e.g. specific knowledge, the level of conceptual understanding (by so-called ordered multiple-choice items) or students’ misconceptions. Similarly, the assessment of e.g. reasoning competence depends on the underlying framework that defines the construct and that might differ between studies. Concerning specific aspects of formative assessment, interesting research approaches and results can be found in the national literature (see D 2.3), e.g. about feedback in Germany or the ‘assessment dialogue’ in Denmark. The instruments listed in the country reports and the respective publications can also be important resources once the decision on specific assessment approaches within the ASSIST-ME project has been made. As the degree of inquiry in instruction might vary from guided to open, formative assessment methods might vary in their level of formality (on-the fly/informal vs. formal) and their time frame (within/between lessons vs. within/between teaching units vs. across semesters/years).

Links between assessment and IBE. In the literature review, no specific patterns linking aspects of inquiry to assessment methods could be observed. Nevertheless, there are aspects that are assessed significantly more often than others which might be either due to the fact that they are considered more important or relevant or that they are regarded as more suitable for assessment. Among the well-researched aspects are e.g. ‘argumentation/reasoning’, ‘communication/debating with peers’ and ‘diagnosing problems/identifying questions’ whereas e.g. ‘researching conjectures’ or ‘constructing and using models’ is much less well investigated. The assessment of inquiry-related competences is heavily dominated by constructed and open response items and, especially in science, multiple choice items. Often standardized achievement tests are used in order to compare the effectiveness of inquiry-based approaches compared to more traditional instructional settings. In these cases, the assessment often does not focus on specific aspects of the inquiry process but on a general construct called ‘inquiry skills’ or ‘science process skills’ for which a precise definition of their meaning is frequently missing. Assessment methods that are widely used in addition to more formal approaches are observational measures such as video, audio or observation notes.

4.2 Recommendations

Taking the work of WP 2 during the first nine month of the ASSIST-ME project and the conclusions presented in the previous section into account, several recommendations might be inferred for the future work in the ASSIST-ME project. These recommendations can be summarized under the following four questions:

1. What should be assessed?
2. How should it be assessed?
3. Which tools or instruments are needed for the assessment?
4. How can teachers be supported in the process?

In the following, recommendations are given with respect to these four questions.

4.2.1 What should be assessed?

Since no generally accepted definition of IBE exists and, as D 2.5 shows, the existing definitions cover a wide range of specific skills, competences, activities and attitudes, it is regarded as crucial for the project to decide on an operational definition of inquiry. It should be emphasized that the aim is not to produce a completely comprehensive definition of IBE but one that is easily accessible by teachers and addresses those aspects of inquiry that will be at the focus of ASSIST-ME.

D 2.5 may serve as the basis for this decision. The operational definition should focus on those competences that the project is going to assess. When defining competences the definition of Weinert (2001) should be considered. It is recommended to choose subject-independent competences such as e.g. using models, investigations or argumentation. To ensure that the chosen competences are really relevant in all three subject domains, experts from the subjects should be included in the decision process. In this context, it is important to keep in mind that for specific competences like e.g. reasoning, different frameworks may exist (e.g. in science sometimes 'defining and working with variables' and the 'need for controlled experiments' is regarded as belonging to reasoning whereas others allocate these competences to planning and conducting investigations). Thus, there is a need of exactly defining what these competences are supposed to mean in the ASSIST-ME context.

In this decision process, however, the content or context in which the competences should be demonstrated and assessed should not be forgotten. In the literature on formative assessment, it is strongly advocated that assessment should not be an afterthought to curriculum development but that curriculum and assessment experts as well as teachers should work together in the development process from the very beginning. This should also be heeded the other way around.

4.2.2 How should it be assessed?

To answer this question, again a need is seen for clear and decisive definitions (e.g. what is meant by methods and formats or instruments and tools and what their consequences are). With respect to methods, one needs to think about e.g. process information (e.g. how to collect data), what features (e.g. reliability, validity …) are expected and how these features can be transformed to make them meaningful for everyone involved in the project including the teachers. In this context, it has to be taken into account that different assessment formats might vary in test quality characteristics like e.g. reliability or validity. More open formats require good rubrics (generic and/or specific); the development of rubrics should be part of the development process from the very beginning. Once the data is collected, one also needs to think about analytical procedures. 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 has to be stressed that communication should be an integral part of the methodology.

It is recommended to select a small number of assessment formats and develop illustrative examples that are close to existing teaching practice as well as guidelines for how to use them (facilitating the later production of teacher training materials). In this context, methods or formats that already exist and are used in the countries (see D 2.3) might provide helpful information and examples of best practice. Examples of methods to assess different aspects of IBE can be found in D 2.4. It is recommended to select methods that focus on different aspects of IBE and that allow for formative assessment at different levels of formality. The decision, thus, should not only be about formats but should also consider competences, processes and analytical procedures. It might be sensible to combine formats that assess specific ('important') competences with formats that assess more the process as a whole (one possible combination might e.g. be multiple choice or open response items for assessing reasoning and notebooks for assessing the whole process of conducting investigations).

Because of their huge prominence in the literature on assessment and IBE, it is recommended to include 'traditional' paper and pencil assessments (multiple choice, constructed or open response items) in the ASSIST-ME assessment formats in order to be compatible to international research. These traditional formats should be combined with innovative formats such as technology-based assessments (for specific recommendations concerning e-assessments see next section). E-assessments have special potential for embedded assessments that can provide almost immediate feedback. These two formats could sensibly be complemented with more overarching, long-term formats like e.g. portfolios or learn logs (which should be used iteratively) and observational measures like videos or interviews to get a deeper understanding of students actions and answers. In this context, it has also to be taken into account that assessment methods are not necessarily universally usable across age groups, grade levels or subjects. Whereas concept maps, e.g., can be useful formats for science in secondary school they might be less usable in the first grades of primary school because children there are still lacking the necessary ability to reflect and abstract. On the other hand, learn logs or learning diaries could be useful tools in primary school, especially in those subjects that have comparably high amounts of instruction time.

All chosen formats, however, should ensure that they involve both, teachers and learners but that ultimately the learner has to take action. They should promote productive discourse and be able to capture the classroom dialogue. This explicitly includes giving good formative feedback on written assignments.

If the assessment is ICT-based, the relationship between e-learning and e-assessment has to be taken into account. The relationship is a key to the success of formative e-assessment. E-assessment loses its impact when it is just an add-on to usual classroom practice and little time and effort is placed on it. It is thus important that the format and functionality of the e-assessments are familiar to the learner. Moreover, to ensure that the outcomes of e-assessment are used validly, the assessment development process should consider which aspects of the curriculum could and do use technology in teaching and learning. The construct being taught and learnt and the construct being assessed need to be linked.

Technology provides opportunities for the blurring of traditional lines between learning, formative and summative assessment like e.g. in gaming models. Using a gaming model shows how formative and summative assessment can be brought together in an IBE environment which develops and recognizes competencies. Moreover, adaptive e-assessment can support the learner with self-assessments.

The immediacy and elaborateness of feedback that can be given to students when it is integrated with e-assessment give it a distinct advantage over other types of assessment. Feedback based on e-assessment should be instant, differentiated, individualised and support students in their learning progress.

Another advantage of e-assessments is the opportunity for interventions and intelligent tutoring options that allow for learner style and preferences to be accounted for, to ensure that e-learning is individualised. In addition, the richness of the outputs from e-assessment allows teachers to monitor individual, group and sub-groups' attainment, to see progress over time and across interventions, all of which feed into a more complete formative education process.

However, there are two factors important for the implementation of e-assessments: the source of impetus for e-assessments and the need for resources and technical support. With respect to the first factor, for e-assessment initiatives to be scaled up, a key factor in success will be the financial commitment and support of senior managers. Some kind of top-down approach is needed as practitioners will not be able to influence change of this nature from the bottom up. Nevertheless, teacher beliefs, skills and confidence are of course also crucial in the implementation process. Moreover, the implementation of e-assessments puts high requirements on financial and technical support (like e.g. interoperability with existing systems, network capacity and technical and pedagogical support for staff).

4.2.3 Which tools or instruments are needed for the assessment?

The answer to this question depends on the type of classroom processes and competences that should be assessed. The project needs formats that allow for looking at outputs of what students are constructing but also for looking at learning processes. 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. They should focus on the learning as it is taking place and provide ways to improvement. In that way, they will inform the implementation of whole learning progressions for developing the competences. Moreover, the content of the tasks should have relevance for everyday life, should challenge the students and must allow for the collection of meaningful data.

Figure 3, Figure 4, and Figure 5 provide useful information in view of more specific methodological recommendations. In general, it is necessary to describe each tool or instrument and its use in detail. By doing this, one has to take the school level into account. For instance, it might be necessary to find alternative tools and instruments for primary school level as most of the reported ones are quite complex.

In science, most of the tools and instruments are mainly used for the assessment of only one aspect of IBE (see Figure 4). For example, discourses and conversations are predominantly used for the assessment of 'argumentation/reasoning/using evidence' or 'debating with peers/communication'. Another example are constructed response items that are also mainly used for the assessment of 'argumentation/reasoning/using evidence'. Thus, it is desirable to develop tools and instruments that can be used for the assessment of more than one aspect of IBE. Preferably, consecutive aspects should be assessed, e.g. diagnosing problems, identifying questions, and formulating hypotheses. In view of an analysis of the inquiry process and related competences, portfolios, mapping techniques and notebooks are possible assessment tools or instruments. Furthermore, the assessment of planning and conducting investigations and experiments should be focused as this aspect forms the core of scientific inquiry. Multiple-choice items are an example for traditional assessment tools or instruments predominantly used for the assessment of knowledge and understanding, respectively. However, they can also be used for the assessment of aspects of IBE, e. g. 'diagnosing problems/identifying questions' or 'formulating hypotheses'.

In technology, on the contrary, it might be necessary to implement tools and instruments that assess single aspects of the engineering design process. Usually, single aspects of IBE are assessed using almost exclusively performance assessments (see Figure 5). The whole process is mostly assessed by portfolios or observations. Here, alternative tools and instruments should be developed, e g. mapping techniques and notebooks.

As well as in technology, in mathematics it might be necessary to develop diverse methods for several purposes. Looking at Figure 3 three areas can be distinguished. First, knowledge and understanding are assessed by several tools and instruments like portfolios, interviews, and observations. A second focus of the assessment is problem solving that is assessed by mostly different methods, e. g. artefacts, discourse, and videos. In contrast in the third area, a certain method is in the focus namely constructed response items. These items are used for assessing a lot of aspects of IBE, e. g. 'argumentation/reasoning/using evidence', 'diagnosing problems/identifying questions', and evaluating results. Moreover, constructed response items are used for the assessment of 'knowledge/understanding' and 'problem solving', thus relating all three areas. Therefore, one objective should be the development of tools or instruments that assess all relevant aspects of IBE in mathematics.

Leading over to the last question, the project needs to 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, especially for more open assessment formats. They can provide concise structures that allow teachers to organize information in a sensible way.

When designing or developing ICT tools or instruments several aspects have to be taken into account. First of all, e-assessments should exploit the affordances of the technology and not just translate a paper-and-pencil test to an online version. This could be achieved by the inclusion of particular elements, for example, scaffolding questions, optional hints and clues, simulations and scenarios or digital learning objects. Moreover, one should make use of the interactivity that e-assessments provide and that could be especially useful in IBE (by interaction with authentic problem solving environments, instant feedback and adaptation to the learners' responses). However, there is a need to avoid 'random button pressing'. Interactive tasks should not be superficial or authoritative, but should give learners control and contribute to deeper, dialogical interaction among students. In addition, aspects of validity and reliability have to be taken into account when developing e-assessments.

4.2.4 How can teachers be supported in the process?

To enable teachers to implement formative assessment approaches in their teaching, they need support. TPD is regarded as vital and indispensable in this respect. Research has shown that this support must not be constrained to content and methods but must also address e.g. teachers' pedagogical beliefs that can be a significant impediment to the implementation of formative assessment. In this context, one also has to think about the vocabulary to make the definitions accessible for teachers. In Germany, teachers e.g. prefer the term 'individual support' to 'formative assessment' – this vocabulary might be to some extend country specific. Following experiences from the UK with respect to 'assessment for learning', it is recommended that the project should focus on teachers experienced in IBE (at least in the beginning). Both, IBE and formative assessment present teachers with a challenge that in combination might be very hard for them to meet. In this context, however, it has to be considered that some countries reported that there is hardly any IBE used by teachers in their daily instruction (e.g. the Czech Republic or Finland). This has to be taken into account when designing the TPD activities. 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. This can be supported by tailoring the assessment formats to ordinary classrooms and designing the assessment tools in a way that they are not regarded as a lot of extra work by the teachers (since time is a serious constraint for teachers).

In view of e-assessments, teachers need support in order to engage and feel confident with the new assessment technology. TPD activities should cover a range of areas: technology use in teaching and learning; e-assessment; formative assessment; links between summative and formative assessment; the pedagogic approaches of IBE and associated competencies; how each of these apply to the STM subjects; and the interrelationships between each of these aspects. Moreover, teachers need to be released from their usual workload in order to take up TPD.

References

Abd El Khalick, F., Boujaoude, S., Duschl, R. A., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., …Taylor, P. C. (2004). Inquiry in Science Education: International Perspectives. *Science Education*, *88*(3), 397–419.

Abi-El-Mona, I., & Abd-El-Khalick, F. (2006). Argumentative Discourse in a High School Chemistry Classroom. *School Science and Mathematics*, *106*(8), 349–361.

ALTA (Ed.). (2009). *Adaptive Learning Teaching Assessment (website)*. Retrieved from www.alta-systems.co.uk

Anderson, R. D. (2002). Reforming Science Teaching: What Research Says About Inquiry. *Journal of Science Teacher Education*, *13*(1), 1–12.

Artigue, M., & Baptist, P. (2012). *Inquiry in Mathematics Education* (Resources for Implementing Inquiry in Science and in Mathematics at School). Retrieved from http://www.fibonacci-project.eu/

Artigue, M., Dillon, J., Harlen, W., & Léna, P. (2012). *Learning through inquiry* (Resources for Implementing Inquiry in Science and in Mathematics at School). Retrieved from http://www.fibonacci-project.eu/resources

Aschbacher, P., & Alonzo, A. (2006). Examining the Utility of Elementary Science Notebooks for Formative Assessment Purposes. *Educational Assessment*, *11*(3 & 4), 179–203.

Ash, D. (2008). Thematic continuities: Talking and thinking about adaptation in a socially complex classroom. *Journal of Research in Science Teaching*, *45*(1), 1–30.

Assessment Reform Group (Ed.). (2002). *Assessment for Learning: 10 principles*. Retrieved from http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf

Barak, M., & Doppelt, Y. (2000). Using portfolios to enhance creative thinking. *Journal of Technology Studies*, *26*(2), 16–24.

Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of Procedure-Based Scoring for Hands-On Science Assessment. *Journal of Educational Measurement*, *29*(1), 1–17.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, *85*(5), 536–553.

Berland, L. K. (2011). Explaining Variation in How Classroom Communities Adapt the Practice of Scientific Argumentation. *Journal of the Learning Sciences*, *20*(4), 625–664.

Bernholt, S., Rönnebeck, S., Ropohl, M., Köller, O., & Parchmann, I. (2013). *Report on current state of the art in formative and summative assessment in IBE in STM – Part I* (No. D 2.4). Kiel.

Björkman, J., Labetzki, T., & Tiemann, R. (2010). Ein Instrument zur Videoanalyse von Scientific Inquiry [An instrument for the video-based analysis of scientific inquiry]. In S. Bernholt (Ed.), *Konzepte fachdidaktischer Strukturierung für den Unterricht. GDCP, Jahrestagung in Oldenburg 2010* (pp. 304–306). Münster: LIT.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Boaler, J. (1998). Open and closed mathematics: student experiences and understandings. *Journal for Research in Mathematics Education*, *29*(1), 41–62.

Brandstädter, K., Harms, U., & Großschedl, J. (2012). Assessing System Thinking Through Different Concept-Mapping Practices. *International Journal of Science Education*, *34*(14), 2147–2170.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*, *57*(2), 213–227.

Brousseau, G., & Balacheff, N. (1997). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970-1990*. Dordrecht: Kluwer Academic Publishers.

Cavagnetto, A., Hand, B. M., & Norton-Meier, L. (2010). The Nature of Elementary Student Science Discourse in the Context of the Science Writing Heuristic Approach. *International Journal of Science Education*, *32*(4), 427–449.

Chin, C., & Teou, L.-Y. (2009). Using Concept Cartoons in Formative Assesment: Scaffolding Students' Argumentation. *International Journal of Science Education*, *31*(10), 1307–1332.

Christensen, T. S. (2004). *Integreret Evaluering – En undersøgelse af den fagligt evaluerende lærer-elevsamtale som evalueringsredskab i Gymnasial Undervisning [Integrated assessment - an investigation of the subject specific assessingly oriented teacher-student dialogue as an assessment tool in the upper secondary education. With an extensive English summary]* (PhD Dissertation). University of Southern Denmark, Odense. Retrieved from http://static.sdu.dk/mediafiles/Files/Om_SDU/Fakulteterne/Humaniora/Phd/afhandlinger/2005/Afhandlinger%2042_spanget%20pdf.pdf

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of Classroom Mathematics Traditions: An Interactional Analysis. *American Educational Research Journal*, *29*(3), 573–604.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a Problem-Centered Second-Grade Mathematics Project. *Journal for Research in Mathematics Education*, *22*(1), 3–29.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. Retrieved from http://igraph.sf.net

Davis, R. S., Ginns, I. S., & McRobbie, C. J. (2002). Elementary School Students' Understandings of Technology Concepts. *Journal of Technology Education*, *14*(1), 35–50.

Dawson, V., & Venville, G. J. (2009). High-School Students' Informal Reasoning and Argumentation about Biotechnology: An Indicator of Scientific Literacy? *International Journal of Science Education*, *31*(11), 1421–1445.

Dolin, J. (2002). *Fysikfaget i forandring – læring og undervisning i fysik i gymnasiet med fokus på dialogiske processer, autenticitet og kompetenceudvikling. [Physics in school under change. Learning and teaching in physics in upper secondary with focus on dialogical processes, autenticity and competence development. With an English summary]* (Phd Dissertation). Roskilde University, Roskilde. Retrieved from http://rudar.ruc.dk/handle/1800/1645

Dolin, J. (2012). *Assess Inquiry in Science, Technology and Mathematics Education: ASSIST-ME proposal.*

Doppelt, Y. (2003). Implementation and assessment of project-based learning in a flexible environment. *International Journal of Technology and Design Education*, *13*(3), 255–272.

Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the matriculation 2000 project. *Journal of Research in Science Teaching*, *40*(1), 34–52.

Fox-Turnbull, W. (2006). The influences of teacher knowledge and authentic formative assessment on student learning in technology education. *International Journal of Technology and Design Education*, *16*(1), 53–77.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Kluwer Academic Publishers.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the Fidelity of Implementing Embedded Formative Assessments and Its Relation to Student Learning. *Applied Measurement in Education*, *21*(4), 360–389.

Furtak, E. M., Shavelson, R. J., Shemwell, J. T., & Figueroa, M. (2012). To teach or not to teach through inquiry: Is that the question? In S. M. Carver & J. Shrager (Eds.), *The journey from child to scientist. Integrating cognitive development and the education sciences* (1st ed., pp. 227–244). Washington, D.C.: American Psychological Association.

Gallin, P. (2012). Dialogic learning – from an educational concept to daily classroom teaching. In P. Baptist & D. Raab (Eds.), *Resources for Implementing Inquiry in Science and in Mathematics at School. Implementing Inquiry in Mathematics Education* (pp. 23–33). Retrieved from http://www.fibonacci-project.eu/resources

Georges, F., & Pansu, P. (2011). Les feedbacks à l'école: un gage de régulation des comportements scolaires. *Revue Française de Pédagogie*, *176*, 101–146.

Goodnough, K., & Long, R. (2006). Mind mapping as a flexible assessment tool. In M. McMahon, P. Simmons, R. Sommers, D. DeBeats, & F. Crawley (Eds.), *Assessment in science: Practical experiences and education research* (pp. 219–228). Arlington: NSTA Press.

Grugeon, B., Pilet, J., Chenevotot, F., & Delozanne, E. (2012). *Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire: Recherches en didactique de mathématiques. Enseignement de l'algèbre, bilan et perspectives* (hors série).

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, *7*, 93–100.

Hammann, M., Phan, T. H., & Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen? [Experimentation as problem-solving: Can the SDDS model be used to measure different dimensions of experimentation?]. *Zeitschrift für Erziehungswissenschaft*, *8*, 33–49.

Harlen, W. (2007). *The Quality of Learning: Assessment Alternatives for Primary Education* (Primary Review Research Survey No. 3/4). Retrieved from http://image.guardian.co.uk/sys-files/Education/documents/2007/11/01/assessment.pdf

Harskamp, E., Ding, N., & Suhre, C. (2008). Group Composition and Its Effect on Female and Male Problem-Solving in Science Education. *Educational Research*, *50*(4), 307–318.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.

Heinze, A., Cheng, Y.-H., Ufer, S., Lin, F.-L., & Reiss, K. (2008). Strategies to foster students' competencies in constructing multi-steps geometric proofs: teaching experiments in Taiwan and Germany. *International Journal of Mathematics Education*, *40*(3), 443–453.

Hickey, D. T., Taasoobshirazi, G., & Cross, D. (2012). Assessment as learning: Enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching*, *49*(10), 1240–1270.

Hickey, D. T., & Zuiker, S. J. (2012). Multilevel Assessment for Discourse, Understanding, and Achievement. *Journal of the Learning Sciences*, *21*(4), 522–582.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, *42*(7), 791–806.

Hong, J.-C., Yu, K.-C., & Chen, M.-Y. (2011). Collaborative learning in technological project design. *International Journal of Technology and Design Education*, *21*(3), 335–347.

International Technology Education Association (Ed.). (2007). *Standards for Technological Literacy: Content for the Study of Technology.* International Technology Education Association. Retrieved from http://www.iteea.org/TAA/PDFs/xstnd.pdf

Jundt, W., & Wälti, B. (2011). *Mathematische Beurteilungsumgebungen*. Bern: Schulverlag Plus.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, *44*(1), 183–203.

Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht [Assessment and modelling of competences within mathematics instruction]. In E. Klieme, D. Leutner & M. Kenk (Eds.), *Zeitschrift für Pädagogik 56. Beiheft. Kompetenzmodellierung. Zwischenbilanz des DFG Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (pp. 64–74). Weinheim, Basel: Beltz.

Klieme, E., & Harks, B. (2013, August). *Formative assessment. General concepts, recent debates in Germany, and findings from experimental studies in mathematics education*, Berlin.

Kubasko, D., Jones, M. G., Tretter, T., & Andre, T. (2008). Is it live or is it memorex? Students' synchronous and asynchronous communication with scientists. *International Journal of Science Education*, *30*(4), 495–514.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet environments for science education*. Mahwah: Lawrence Erlbaum Associates Publishers.

Looney, J. W. (2011). *Integrating Formative and Summative Assessment: Progress Toward a Seamless System?* (OECD Education Working Papers No. 58). Retrieved from http://www.oecd-ilibrary.org/education/integrating-formative-and-summative-assessment_5kghx3kbl734-en

Mason, L. (2001). Introducing talk and writing for conceptual change: a classroom study. *Learning and Instruction*, *11*(4-5), 305–329.

McLoughlin, E. (2011). *Guide for developing ESTABLISH teaching and learning units.*

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*(2), 233–268.

Ministry of Children and Education (Ed.). (no date). *Værktøjer til evaluering i folkeskolen*. Retrieved from http://uvm.dk/Uddannelser-og-dagtilbud/Folkeskolen/De-nationale-test-og-evaluering/Evaluering/Vaerktoejer

Nantawanit, N., Panijpan, B., & Ruenwongsa, P. (2012). Promoting Students' Conceptual Understanding of Plant Defense Responses Using the Fighting Plant Learning Unit (FPLU). *International Journal of Science and Mathematics Education*, *10*(4), 827–864.

National Research Council (Ed.). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, D.C.: The National Academies Press.

Nielsen, J. A. (2012). Co-opting Science: A preliminary study of how students invoke science in value-laden discussions. *International Journal of Science Education*, *34*(2), 275–299.

Nohda, N. (1995). Teaching and Evaluating Using "Open-Ended Problems" in Classroom. *Zentralblatt für Didaktik der Mathematik*, *27*(2), 57–61.

Nohda, N. (2000). Teaching by Open-Approach Method in Japanese Mathematics Classroom. *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education (PME)*, (1), 39–53.

OECD (Ed.). (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: Organisation for Economic Co-operation and Development.

OECD (Ed.). (2013). *PISA 2015: Draft Science Framework*. Retrieved from http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm

OECD & Centre for Educational Research and Innovation (Eds.). (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: Organisation for Economic Co-operation and Development.

Office for Standards in Education, Children's Services and Skills (Ofsted) (Ed.). (2011). *The impact of the 'Assessing pupils' progress' initiative*. Retrieved from http://www.ofsted.gov.uk/resources/impact-of-assessing-pupils-progress-initiative

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the Quality of Argumentation in School Science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.

Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the Integration of Technology and Assessment. *Journal of Research on Technology in Education*, *43*(2), 119–134.

Pijls, M., Dekker, R., & van Hout-Wolters, B. (2007). Reconstruction of a collaborative mathematical learning process. *Educational Studies in Mathematics*, *65*(3), 309–329.

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C., …Foley, B. (2006). Fifth Graders' Science Inquiry Abilities: A Comparative Study of Students in Hands-On and Textbook Curricula. *Journal of Research in Science Teaching*, *43*(5), 467–484.

Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

Polya, G. (1957). *How to solve it* (2nd ed.). Princeton, NJ: Princeton University Press.

R Core Team (Ed.). (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, *27*, 63–73.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science Education Now: A Renewed Pedagogy for the Future of Europe*. Luxemburg: Office for Official Publications of the European Communities.

Rönnebeck, S., Bernholt, S., Ropohl, M., Köller, O., & Parchmann, I. (2013). *National reports of partner countries reviewing research on formative and summative assessment in their countries* (No. D 2.3). Kiel.

Rönnebeck, S., & Köller, O. (2013a). *Guidelines for partner countries to facilitate collecting national research on assessment* (No. D 2.1). Kiel.

Rönnebeck, S., & Köller, O. (2013b). *Synopsis of the literature review* (No. D 2.2). Kiel.

Rönnebeck, S., Ropohl, M., & Köller, O. (2013). *Report of outcomes of the expert workshop on formative assessment in STM and IBE* (No. D 2.6). Kiel.

Ropohl, M., Rönnebeck, S., Bernholt, S., & Köller, O. (2013). *A definition of inquiry-based STM education and tools for measuring the degree of IBE* (No. D 2.5). Kiel.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal Formative Assessment and Scientific Inquiry: Exploring Teachers' Practices and Student Learning. *Educational Assessment*, *11*(3-4), 205–235.

Ruiz-Primo, M. A., Li, M., Ayala, C., & Shavelson, R. J. (2004). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education*, *26*(12), 1477–1506.

Ryu, S., & Sandoval, W. A. (2012). Improvements to Elementary Children's Epistemic Understanding from Sustained Argumentation. *Science Education*, *96*(3), 488–526.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Sampson, V., Grooms, J., & Walker, J. P. (2011). Argument-Driven Inquiry as a Way to Help Students Learn How to Participate in Scientific Argumentation and Craft Written Arguments: An Exploratory Study. *Science Education*, *95*(2), 217–257.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. San Diego: Academic Press.

Schoenfeld, A. H. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning. A Project of the National Council of Teachers of Mathematics* (pp. 334–370). New York: MacMillan Publishing Company.

Schoenfeld, A. H., & Kilpatrick, J. (2013). A US perspective on the implementation of inquiry-based learning in mathematics. *ZDM - The International Journal on Mathematics Education*. Advance online publication. http://link.springer.com/journal/11858

Scholar (Ed.). (no date). *SCHOLAR in Scotland at the school/university interface*. Retrieved from www.scholar.hw.ac.uk

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students’ task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics*, *79*(2), 215–237.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance Assessment in Science. *Applied Measurement in Education*, *4*(4), 347–362.

Shemwell, J. T., & Furtak, E. M. (2010). Science Classroom Discussion as Scientific Argumentation: A Study of Conceptually Rich (and Poor) Student Talk. *Educational Assessment*, *15*(3), 222–250.

Smit, R., & Birri, T. (2012). *Lernen mit Rubrics als Teil der formativen, standardorientierten Beurteilung.*, PH St. Gallen.

Socrative.com (Ed.). (2013). *Socrative. Smart Student Response System*. Retrieved from http://www.socrative.com/how-it-works.php

Talbot, L. (2009). *L’évaluation formative : comment évaluer pour remédier aux difficultés d’apprentissage*. Paris: Armand Colin.

Thomson Reuters. (2013). *Web of knowledge – Journal citation reports*. Retrieved from http://admin-apps.webofknowledge.com/JCR/JCR?PointOfEntry=Home&SID=3F36 dpJCK emKLP7aK2p

Valanides, N., & Angeli, C. (2008). Distributed Cognition in a Sixth-Grade Classroom: An Attempt to Overcome Alternative Conceptions about Light and Color. *Journal of Research on Technology in Education*, *40*(3), 309–336.

Vellom, R. P., & Anderson, C. W. (1999). Reasoning about data in middle school science. *Journal of Research in Science Teaching*, *36*(2), 179–199.

Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–65). Göttingen: Hogrefe & Huber.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, *16*(1), 3–118.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, *11*(3-4), 283–289.

Wilson, M., & Sloane, K. (2000). From Principles to Practice: An Embedded Assessment System. *Applied Measurement in Education*, *13*(2), 181–208.

Winter, F. (2007). Fragen der Leistungsbewertung beim Lerntagebuch und Portfolio [Questions concerning the evaluation of learning diaries and portfolios]. In M. Gläser-Zikuda (Ed.), *Lernprozesse dokumentieren, reflektieren und beurteilen: Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* (pp. 109–129). Bad Heilbrunn: Klinkhardt.

Wylie, E. C., & Dolan, R. P. (2013). *The Role of Formalized Tools in Formative Assessment: Paper presented at the Annual Meeting of the American Educational Research Association*, San Francisco.

Yiping, L., & Blanchard, P. (2007). *National and state science as inquiry standards*. Retrieved from http://pti.lsu.edu/standards.aspx

Figures

Figure 1: Different types of formative assessment (Klieme & Harks, 2013)................10
Figure 2: Formative assessment cycle (Klieme & Harks, 2013) 11
Figure 3: Interrelations between aspects of IBE (red) and assessment methods (blue) in mathematics education (based on the reviewed articles) 24
Figure 4: Interrelations between aspects of IBE (red) and assessment methods (blue) in science education (based on the reviewed articles) 25
Figure 5: Interrelations between aspects of IBE (red) and assessment methods (blue) in technology education (based on the reviewed articles) 26

Tables

Table 1: Deliverables from Work Package 2 5
Table 2: Keywords and topics for the search in databases in science education.........13
Table 3: Relevant journals and their impact factors..13
Table 4: Final extract for the literature review 14
Table 5: Distribution of reviewed articles across the identified aspects of inquiry........16
Table 6: Aspects of IBE that are incorporated in at least 10% of the research papers within each of the three domains 18