The work in the 7 Work Packages (WP) resulted in a total of 44 deliverables. This part of the final report will be organised after these deliverables in so far as they report results discovered during the project period.

Work Package 1 - *Project Management*

Deliverables:
D1.1 Consortium agreement
D1.2 Semi-annual Status Reports
D1.3 Public ASSIST-ME Website
D1.4 Private Project Website
D1.5 Periodic and Final Reports

No results *per se*

Work Package 2 - *Synthesize Existing Research*

Deliverables:
D2.1 Guidelines for partner countries to facilitate collecting national research on assessment
D2.2 Synopsis of the literature review.
D2.3 National reports of partner countries reviewing research on formative and summative assessment in their countries
D2.4 Report on current state of the art in formative and summative assessment in IBE in STM – Part I and part II
D2.5 A definition of inquiry-based STM education and tools for measuring the degree of IBE
D2.6 Report of outcomes of the expert workshop on formative assessment in STM and IBE
D2.7 Recommendation report

Objectives (from the proposal):
- To review what we know about formative and summative assessment of competences in STM and what methods can work to improve student outcomes, including use of ICT tools.
- To collect recent work that has been conducted in the partner countries on assessment in STM.
- To synthesize the outcomes from other EU-projects on IBE, e.g., Mind The Gap, S-TEAM, ES-TABLISH, Fibonacci, INQUIRE, Primas, SAILS.
- To formulate an operational definition of IBE related to STM and give guidelines for measuring the degree of IBE.
- To provide recommendations for WP4 and WP5.

Selected Results:
"In the first step of the project, a literature review was conducted in order to gather information about the current state of the art in formative and summative assessment in inquiry-based education (IBE) in STM. Searches were conducted in data bases, in the most important journals in the field of STM education, and in the reference lists of relevant publications. […] Especially in science education, numerous publications were found by the search strategies whereas in technology and mathematics education the numbers of publications are much lower. On the one hand, the chosen keywords and search strategies might be a reason. On the other hand, the research foci of the disciplines might be another reason. The results of the literature review indicate that only a small

number of empirical studies have simultaneously investigated both the use of formative and summative assessment in the learning of inquiry in STM and the influence of this form of assessment on the learning of inquiry in STM. Moreover, most of the studies did not assess inquiry directly, but rather knowledge, understanding or attitudes. Nevertheless, there are examples of methodological approaches which illustrate the successful application of several assessment instruments and explain their advantages or disadvantages" (D2.4, p. 4).

It was quite clear that performing the literature review was not a straightforward task. "One major challenge for the literature review was that the field of interest is not clearly defined. With respect to science education, there is still disagreement among researchers and educators about what features define the instructional approach of IBE […] A rich vocabulary is used to describe inquiry-based approaches to teaching and learning, such as inquiry-based teaching and learning, authentic inquiry, model-based inquiry, modelling and argumentation, project-based science, hands-on science, and constructivist science […] These approaches might include characteristics of IBE to a varying degree but they are not necessarily synonyms of IBE. The situation gets even more complicated because, e.g. in the US, the field of science education has moved away from using the term inquiry and now [refers to 'practices'] […] Moreover, the definitions of IBE or inquiry-based approaches to teaching and learning differ between the three domains of science, technology, and mathematics (see D 2.5)" (D2.4, p. 5).

"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. […] 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 lev- el of formality (on-the fly/informal vs. formal) and their time frame (within/between les- sons vs. within/between teaching units vs. across semesters/years). […] 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." (D2.7, p. 33-34).

"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." (D2.7, p. 29)

Work Package 3 - *Characterize Educational Systems*

Deliverables:

D3.1 Educational system factors influencing student assessment methods in science, technology and mathematics education

D3.2 National Reports: Educational system factors influencing student assessment methods in science, technology and mathematics education

D3.3 Matrix of educational system factors influencing student assessment methods in science, technology and mathematics education

D3.4 Educational system factors influencing student assessment methods in science, technology and mathematics education

Objectives (from the proposal):

- To identify educational system variables with relevance to formative or summative assessment and IBE in STM
- To map out the participating countries with respect to these variables.
- To provide the consortium and especially WP4, 5 and 6 with systemic information needed for trial implementations.

Results:

D3.4 provides "systemic information needed for trial implementations of formative and summative assessment (FA/SA) methods in inquiry based science, technology and mathematics education (IBSTME). It consists in mapping out the partners' countries following five dimensions that results from a previous survey through the consortium (deliverable 3.1). That led to the elaboration of an online questionnaire which was responded by each partner; each dimension is filled out by a set of close-ended questions and one open- ended question (deliverable 3.2). A first report presenting the results of the online questionnaire was delivered to each member in order to be commented by some national experts (deliverable 3.3). Thus, the characterization and the comparison of the educational systems results from a quantitative and qualitative methods. The quantitative approach relies on close-ended questions, which allows us to compare the different educational systems on the same basis. The researchers' answers to the questionnaire was analysed using a quantitative method. This is based on multiple correspondence analysis (MCA) with countries as subjects and questions from one dimension at a time as variables. The qualitative approach consists on analyzing the open-ended questions and the comments from the experts in order to grasp the fine details of each educational system and to moderate the results from the close-ended questions. The outcomes of this study are twofold. A first part is related to the National Stakeholders Panel (NSP), and a second to the Teacher Expert Panel (TEP) and the Local Working Groups (LWG). Regarding to the NSP, two

ways to reinforce the understanding of ASSIST-ME methods and goals appear depending on the system organization and management:

- Centralized countries need to interact mainly with the central deciders: Cyprus (CYP), France (FR) and Germany (GER) if we only take into account the Schleswig-Holstein German State of the ASSIST-ME partner.
- Decentralized countries need to interact manly with schools and teachers. This is the case for United-Kingdom (UK), Denmark (DK) and Finland (FIN)
- Two countries move between these two poles depending of variables and criteria. This is the case for Switzerland (SW) and Czech Republic (CZ).

The uptake of IBSTME and FA appears as strongly external to teacher culture in any countries. Nevertheless, regarding in-service teachers, except for CYP, these two subjects are part of some CPD programs since 5 years at least. Thus, except for CYP, the members need to reflect with NSP about two questions:

- How to renew the CPD programs towards IBSTME and FA/SA in order to emphasize the interest of the combination of these two approaches? The teacher population is quite experienced and this entails that CPD programs need to be innovative.
- Is it necessary to enhance the part of IBST and FA/SA in CPD programs?

The uptake of the competence model is very diverse. Some members cannot rely on any habit or skills in the use of a competence model by teachers (DK, SW and UK). Some cannot rely on the uptake of the competence model by secondary schools teachers (CYP, CZ, DK and FR). Most of members need to emphasize this uptake in science education (GER, SW) and others in mathematics education (CYP, DK, FR). In UK the competence model isn't very developed, thus the linkage with science and mathematics education is very poor. In CZ, the situation is opposite but as in other countries the teachers are not very trained to use this model. We suggest that each NSP reflects on this specific point –the role of competences in education– in order to orientate the final dissemination process towards the best targets. […] In order to promote FA/SA methods within IBSTME, ASSIST-ME may rely on:

- Teachers' habit to collaborate in most of the countries.
- Monitoring of students' learning at a school level (CYP, CZ, DK, FR, UK).
- Heads of school's responsibility in improving educational outcomes and in orientating teachers towards relevant CPD programs in most countries.

Thus, we suggest supporting this collaboration through the design and use of FA/SA methods that will be designed by ASSIST ME. On another hand, we suggest that specific methods were designing for primary teachers since science is an integrated subject at this level in most countries. This design need also to take into account that mathematics is a specific subject, separated from science and technology, in all countries and for all levels. Developing day-to-day assessment, as part of formative assessment, is a crucial point in all countries. In some countries teachers may rely on resources that already exist (CYP and DK); in this case, we suggest minding the alignment of these actual resources with the formative assessment methods that will be produced by the ASSIST-ME project. In the other countries teachers are not used to rely on this kind of resources. Thus, the project needs to carefully foresee how to facilitate the uptake of the resources by teachers. Summative assessment is a crucial matter in all countries. In some countries teachers have a role in the design of summative assessment (FIN, FR, SW and UK) thus the resources that will be produced by the project might be useful for them. In the other countries where the local authorities have a role in the design of summative assessment, exchanges with each NSP are required in order to connect the ASSIST-ME methods with the national rules" (D3.4, p. 3-5)

Work Package 4 - *Design Assessment Methods*

Deliverables:
D4.1 First list of formative assessment methods
D4.2 Partner feedback to first list
D4.3 Revised version of list, matrix, framework
D4.4 Manual for Teacher Expert Panels
D4.5 Established teacher expert panels in all partner countries
D4.6 National reports after expert panels
D4.7 Description of the ASSIST-ME assessment methods and competencies (Practical examples)

Objectives (from the proposal):
- To determine and finalize a list of formative assessment methods that will be tested.
- To describe ways to combine these formative methods with summative assessment methods
- To determine the competences that are the outcome of the inquiry-based lessons.
- To map out a matrix which relates the assessment methods to domain specific and transversal competences.
- To develop an assessment protocol, i.e. specific guidelines for each assessment method and how to monitor it in daily instruction

Results:
"One of the goals of ASSIST-ME work package 4 was to describe formative assessment methods and competences related to inquiry learning that should be assessed by these assessment methods. The final version of the assessment methods and the competences can be found in [D2.4]". (D4.7, p. 3)

Based on the work in WP2 and WP4, *inquiry teaching* was defined in the following ways:
- "In science education, the process of scientific inquiry involves developing an understanding of scientific aspects of the world around through identifying questions, searching for relevant information, formulating hypotheses or making predictions, planning and carrying out experiments, analyzing, interpreting and evaluating data and results, developing explanations, constructing and using models, engaging in argumentation from evidence and being able to communicate scientifically in different situations and at all steps of the inquiry process." (D4.7, p. 5)
- "In technology education, the process of engineering design involves developing an understanding of the nature and principles of design and technology through defining and delimiting a problem, need or desire, identifying constraints and criteria, investigating relevant information, generating and evaluating possible solutions, analysing alternatives, selecting a potential solution, justifying the decision, planning design of prototype, constructing prototype (using suitable tools/materials), testing prototype by collecting, analysing, interpreting and representing data, evaluate prototype against the criteria, reasoning, modifying the design and redesigning if necessary and communicating at all stages of the process." (D4.7, p. 5)
- "In mathematics education, the process of problem solving involves developing an understanding of mathematics through describing and understanding mathematical or ‚real world' problems, transferring problems into the ‚mathematical world' (if necessary), exploring problems and making conjectures, identifying what is known and what is unknown, creating, using and manipulating mathematical representations, planning and carrying out a problem solving strategy, making and analyzing connections, evaluating the strategy, the conjectures

and the meaningfulness of the results, generalizing and systematizing the results and the specific problem solving strategy and communicating one‘s actions by using adequate vocabulary/representations at every stage of the process.” (D4.7, p. 5)

Based on the work in WP2 and WP4, *formative and summative assessment* was defined in the following ways:

- “Formative assessment is ‘assessment for learning‘ that aims at supporting and improving student learning as it takes place. It has to be designed to support all of the phases of inquiry-based learning as set out in the explanations given above. It is classroom-based, can be at individual or group level as well as criterion referenced. It has a procedural character and can be implemented in many ways: as students conduct activities, they may be helped by immediate question and suggestions to challenge their thinking. On a longer time-scale, data on the achievements made may be collected and interpreted in relation to achievement goals and assessed in the light of student-based criteria. Such data may both indicate the feedback which could be given to help students, individually or collectively, to enhance the learning already achieved, and may also be used to inform decisions about how to reach the next learning steps through the design of new student activities. At all stages there is feed- back from teacher to student and from student to teacher (and from student to student into the teaching-learning process). Formative assessment thus has the function of individual and group support but at the same time gives students an active role in all steps of the assessment process.” (D4.7, p. 6)
- “Summative assessment is ‘assessment of learning‘ with the goal of summarizing, evaluating and reporting learning at a particular point in time. Data is collected and interpreted related to predefined achievement goals (either as a survey or following a student activity) and a judgement is made which is criterion referenced using the same criteria for all students to ensure the comparability of achievement results. The data may also be shared with students to help them develop the ability to review their own work. The process results in a judgment and report of what the student has achieved at a particular time in relation to the goals or standards.” (D4.7, p. 6)

Further, “[w]ithin the ASSIST-ME project, the following three criteria to characterize formative assessment were chosen: 1. Active student involvement in the whole assessment process. 2. The criteria used in the judgment of student activity are student-referenced as well as subject-specific (meaning that the feedback to the student is adapted to the individual student). 3. The immediate aim of the assessment is the identification of further activities that promote the individual students’ learning.” (D4.7, p. 7)

A core part of WP4 consisted of identifying and describing the assessment methods, to be trialled in phase 2 of the project.

“A feedback method is a way of (1) diagnosing the student's level of achievement related to criteria and (2) to give feedback and to start a dialogue with the student(s) in regard of his / her learning progress and of the next steps in learning to be taken. This interaction could take place between the teacher and the student(s) or between peers. The concrete use of the different feedback methods is illustrated by paradigmatic examples. The following paragraphs will introduce four feedback methods. For each of those, the basic principle will be described followed by small varieties. The four feedback methods include:

1) Interactions on-the-fly which describes informal formative assessment of individual students or small groups of students. On-the-fly assessment cannot be planned be- forehand but takes place spontaneously when the teacher recognises good opportunities.
2) Marking (grading and written comments) which describes assessment and feed- back on - often written - pieces of student work. Contrary to the aforementioned method, marking does not take place spontaneously but in planned situations.
3) Self- and peer-feedback describe reflections on the students' own learning as well as assessment which is provided by peer students. In both varieties, it is most important that the students clearly understand the criteria of success.
4) Open classroom discussion and structured classroom dialogue summarizes different varieties of oral assessment.

These four feedback methods reflect the diversity of formative assessment." (D4.7, p. 18)

"Based on the model in Bernholt et al., 2013; Ropohl et al., 2013 and Rönnebeck et al., 2013, specific competences relevant in inquiry were selected for assessment in ASSIST-ME workpackage 4.

- Investigation (in science education)
- Problem solving (in Mathematics education)
- Engineering design (in Technology education)
- Argumentation
- Modelling
- Innovation

The competences are not completely separable but that is not considered as a major problem: teachers have a choice of different ways to assess their students' inquiry-based learning anyways. In order to draw the attention away from the sole "conducting experiments" towards other IBE-relevant competences and activities, it might even be good if there is a certain overlapping of competences assessed.
Each of the six competences will be further refined into sub-competences. The sub- competences with associated learning progressions can be found in chapters 8.2 - 8.7 [of D4.7]." (D4.7, p. 33)

Work Package 5 - *Trial Implementation of Assessment Methods*

Deliverables:
D5.1 Trial implementation plan
D5.2 A working plan for LWG and for the Cross-country steering Committee
D5.3 A teacher profile questionnaire
D5.4 Plan and hold workshops for project partners
D5.5 Teacher self-reporting tools for each assessment method
D5.6 A researcher reporting method for each assessment method
D5.7 Guidelines for production of classroom video study
D5.8 Summary of trial results round one
D5.9 Summary of trial results round two
D5.10 Summary of trial results round three
D5.11 Synthesis of factors

Objectives (from the proposal):

- To steer and coordinate the research on the trial implementation processes of the formative assessment methods from WP4 on the basis of WP2 and WP3 outcomes.

- To describe a frame for the trial implementations of the assessment methods and to plan a distribution of the methods among the partners securing coverage of all variables for all subjects and educational levels.
- To decide the methods used for research of the trial implementations of the assessments.
- To collect data from all partners and sum up the results in a synthesis.

Results:
"The following figure depicts the workflow associated with the trial implementations and illustrates how these are to be used for research purposes but also for refining the assessment methods. It is important to note the iterative nature of the implementations (each of the three trial implementations to be undertaken by each participating country involves three rounds of enactment-data collection-redesign/refinement)." (D5.2, p. 5)

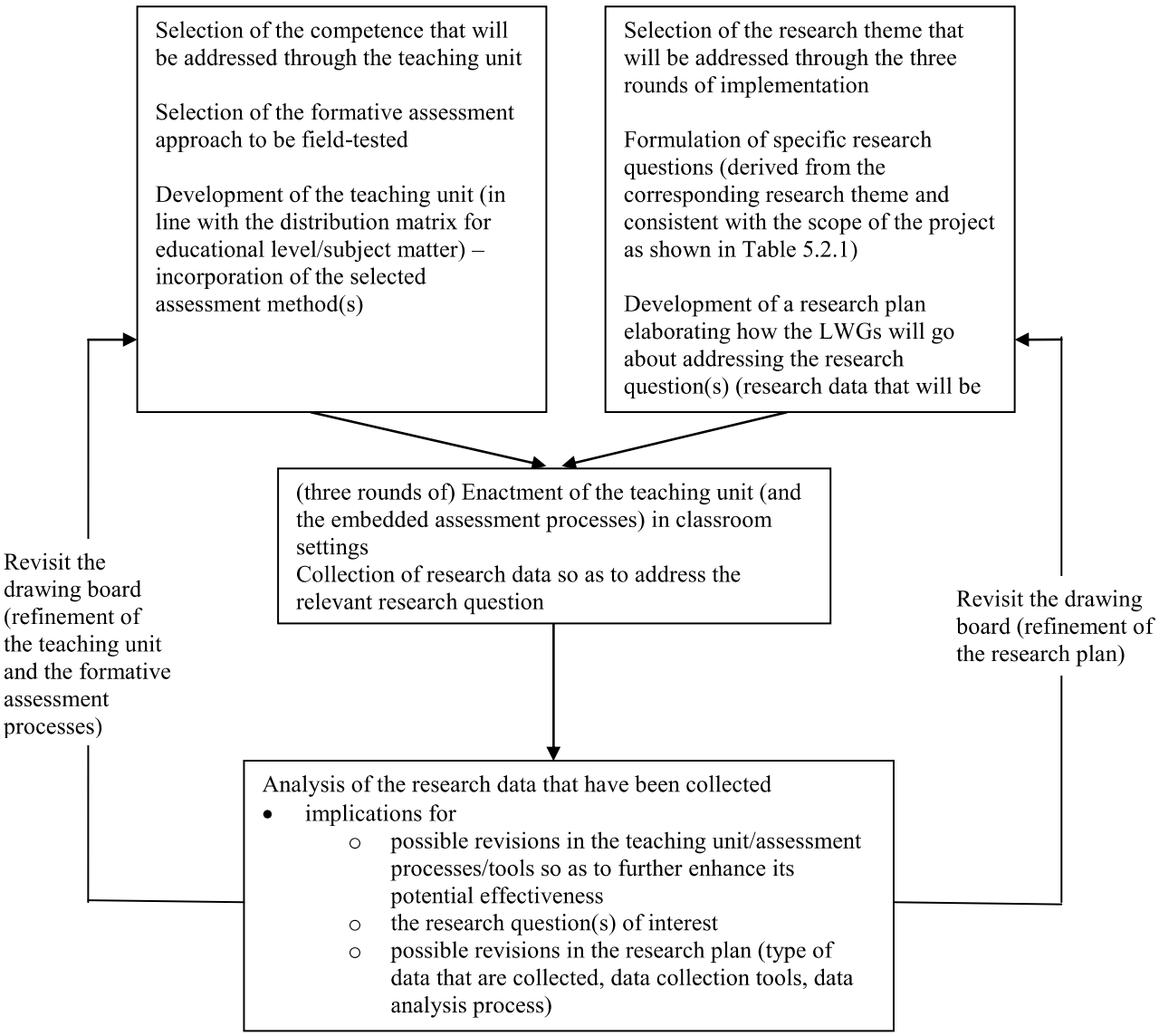

Figure 5.2.1: Workflow for classroom evaluation and refinement of an assessment method by each LWG

A total of 144 implementations of assessment methods were conducted in the project. The following table from D5.11 (p. 5) illustrates the distribution of these implementations.

Implementations (all three rounds of iteration)

Assessment method	Marking (Grading and Written Comments)	Self and Peer Feedback	Interactions on the fly	Structured assessment dialogue
Implementations	33	34	40	37
Total number of students	798	1008	973	approx. 750
Total number of teachers	29	32	24	15

Findings related to written feedback

"Overall, the feedback comments of all the rounds appeared to be rather thorough in terms of coverage of the relevant criteria that were specified by the teachers and discussed with the students. These criteria were reflecting the key aspects of the competence of interest. The mean value for this variable and the median indicate that in their majority the teachers offered comprehensive comments that addressed most of the specified criteria ($\bar{x}$= 4.42, Mdn=5, SD=0.783). […] A possible interpretation for this finding pertains to the benefits emerging from the establishment and operation of LWGs. In particular, this may have provided a productive structure that served to support teachers in enacting this assessment method. For example, one part of this procedure involved the articulation of the important aspects of the given competence in each case and the discussion of how these were to be negotiated with the students in the class. Another interpretation relates to the use of specially designed tools, intended to systematize teachers' effort to address the relevant aspects of the competence of interest. […]. Such tools could have helped maintaining a high degree of alignment between the focus of the feedback and the essential aspects of the competence of interest […] The analysis of the data in view of the second component, namely the provision of justification to the students about the assessment, involved two stages. Firstly, we examined the references made by the teachers to what the students had achieved, or had not achieved, with respect to the specified criteria. […] references to what the students had achieved appeared in 87% whereas references to what they had yet to achieve occurred in 70% of the cases […]In the second stage, we distinguished those references into two categories depending on whether they were accompanied by a relevant justification addressed to the student. A reference was considered justified in cases when the teacher provided reasons to the students, as to how s/he knew that the student had, or had not, achieved a certain criterion (or, put differently, had, or had not, attained a specific aspect of the competence of interest). […] The justified references to what the students had not achieved ($\bar{x}$= 3.20, Mdn=3, SD=1.149) were more frequent than the justified references to what they had achieved ($\bar{x}$= 1.79, Mdn=1, SD=1.408). […] Interestingly, even though the teachers referred to what the student has achieved more frequently (87%) than to what they had yet to achieve (70%), they happened to provide more justified references about students' weaknesses ($\bar{x}$= 3.20, Mdn=3, SD=1.149) than students' strengths ($\bar{x}$=1.79, Mdn=1, SD=1.408). One possible reason for this seeming tendency of the teachers to offer more justifications for their references to what their students had not yet accomplished, could be related to their belief that the students could benefit more by those justifications since they might be providing students with an indication as to what they need to improve. […] Another aspect of the feedback that was coded, by

means of the Likert scale items included in the coding tool, focuses on the guidance provided by the teacher to their students as to how to go about improving their artefact and, hence, strengthening underdeveloped aspects of the competence. The mean score overall (for all three rounds) that was assigned to the teachers' feedback comments in terms of the specificity of the guidance they incorporated was 2.70 (out of a maximum value of 4), with a median value of 3 and a standard deviation of 0.781. […]One point that warrants mention relates to the connection between the analysis for the first and the third aspect of the template (i.e., extent to which the teachers' feedback comments took into account the relevant aspects of the competence and provision of guidance). In an attempt to investigate if there is a statistically significant difference between these two variables we employed Wilcoxon signed-rank test. The results seem to be suggesting that teacher's feedback was coded consistently higher on the first variable (extent to which their feedback comments took into account the relevant aspects of the competence) compared to the second (provision of guidance). In view of the relatively high mean, associated with the first variable this finding seems to be suggesting that whereas the teachers tended to provide feedback comments that took into account the relevant criteria, they found it rather difficult to offer specific guidance on concrete next steps ($p < .001$, $z = -15.277$, $r = -.76$). This result should be treated with caution due to the limited sample size. This limitation notwithstanding, we take the perspective that this finding is hardly surprising. In particular, the provision of guidance can be much more creative, but also demanding, in the sense that it deals with aspects that cannot be anticipated beforehand and need to be judged by the individual teachers in a case-by-case manner, depending on their judgment of the content of the artefact and the personal characteristics of the corresponding students. […] In any case, one claim that could be made in view of these results, is that there seems to be a need to support teachers in performing this task more effectively. In an attempt to address this, Cyprus in the second implementation sought to develop a specially designed tool which contained a large sample of illustrative examples on which the teacher could draw upon in producing their own written feedback comments […] The idea behind this tool was that it could support teachers' attempt to formulate useful feedback. Nevertheless, this hypothesis was not confirmed by the data that emerged in that particular case. Specifically, in many cases (in 34 out of the 46 cases, 74%), the teacher happened to offer feedback comments that were not well aligned with the students' needs […] An extreme example that we encountered in the data reported by the Cyprus LWG, which could serve to illustrate this unproductive use of the tool involves asking students to include additional features in his/her artefact even though these were already there. One possible interpretation for this seemingly unproductive use of the tool is that the teacher found it convenient to rely on the given paradigmatic examples of feedback that were incorporated in the tool, often not in a thoughtful manner. The key idea that emerged from this is that whereas teachers can benefit substantially from tools intended to scaffold the provision of feedback comments, this should not be taken for granted. For instance, it could be argued that the given examples seemed to have subtracted from the teacher's responsibility to interpret students' products, diagnose needs and reflect on possible useful feedback. Hence, exposing teachers to examples of feedback doesn't always guarantee that teachers will effectively use them to promote their students' learning outcomes. In the concluding part of the report we revisit this claim and we elaborate on its implications. […] The last component that was addressed in the analysis of the teachers' feedback relates to their affective side. For this we examined the number of the elements of the feedback comments that carried affective connotations (e.g. "Bravo! You correctly identified the independent variable and specified a procedure that allows changing the independent variable") and we then investigated whether these connotations were mostly positive or negative. It could be argued that the teachers did not appear to be very attentive to the affective side since their feedback received a relatively low score, overall, on this dimension ($\bar{x}$=1.81, Mdn=2, SD=0.802). […] It is also

important to note that in the cases where the feedback comments did contain affective connotations, these were overall rather balanced with a tendency towards the positive side ($\bar{x}$=3.38, Mdn=4, SD=1.198)." (D5.11, p. 13-22)

Findings related to peer-feedback:
"The first component that was addressed relates to the affective side of the feedback comments. For this, we examined the number of the elements of the feedback comments that carried affective connotations and we then investigated whether these were mostly positive or negative. One example of positive affective connotations from one LWG from the Czech Republic is the following: "It is explained gorgeously and written nicely as well". Overall, the feedback comments appeared to be balanced ($\bar{x}$=3.22, Mdn=3, SD=0.882). […] [The component] Alignment of feedback comments with the targeted aspects of the competence […] focuses on the extent to which students took into account the specified criteria while providing feedback to their peers and whether they drew on those in a thorough manner. Most of the students' feedback comments included all the criteria ($\bar{x}$=3.61, Mdn=4, SD=0.714). […] An interpretation for this finding is that the use of the various tools, which contained the intended criteria associated with the competence of interest, helped students recognize the aspects of the relevant competence in their peers' artefacts and then to provide feedback based on them. That is, the use of a structured assessment form, based on the given criteria in each case, made it easier for the students to offer feedback comments anchored on the relevant criteria. The provision of assessment criteria increases the probability of establishing a common understanding between students and the teacher of what has to be assessed (Orsmond, Merry, & Reiling, 2000). […] One point that warrants mention relates to one of the implementations held in Cyprus in the first round, where the assessors were not provided with any actual scaffold other than a simple instruction on how to evaluate their peers' artefacts which did not incorporate any specified criteria or structure for creating comments. In this case, the assessors' comments did not capitalize on the relevant criteria. Rather, they emphasized criteria that were neither directly related to the competence of interest nor powerful in terms of helping their peers improve on the relevant competence. Indicatively, these criteria pertained to superficial aspects of the artefact (e.g. use of colours) and grammar or spelling issues. An additional interesting aspect of the data that emerged in this implementation comes from the interviews with the students who explicitly acknowledged the difficulties they encountered in their attempt to provide comments […] Even though, as mentioned above, in most cases (except for the implementation that did not include any scaffolding) the students' comments considered the pre-specified criteria, they did not exhibit the highest rigor ($\bar{x}$=2.46, Mdn=2, SD=0.903). […] The third component of the coding tool focuses on the extent to which the peer-assessors drew on the intended criteria while constructing their own artefacts. The data suggested that the peer-assessor(s) constructed his/her (their) own artefact by addressing in most cases all the intended criteria ($\bar{x}$=3.29, Mdn=3, SD=0.809). […] Additionally, when constructing their own artefact, the majority of the assessors embodied, in a seemingly adequate manner, each criterion in a thorough manner ($\bar{x}$=2.81, Mdn=3, SD=0.815). This implies that students did seem to appreciate how the relevant criteria could be integrated in the artefact that they were asked to develop. […] The fourth component concerns the justifications provided by the assessor about what the assessee has (or has not) achieved. A feedback comment was considered justified when the assessor provided the assessee with reasons for the credit s/he has given him/her or for the possible weaknesses s/he has alerted him/her for. The justifications of what the students had already achieved for all the three rounds ($\bar{x}$=2.68, Mdn=3, SD=1.392) are approximately equal to the cases of the justifications about what the student had not yet achieved ($\bar{x}$=2.61, Mdn=2, SD=1.436). […] The final component of the reporting template focuses on the guidance provided by the assessor(s) to the assessee(s). It relates to (i) the extent to which the feedback was specific

enough, suggesting concrete next steps that could be taken for improving their artefact and (ii) the extent to which the guidance is indeed related to the competence of interest. Specifically, the guidance provided by the majority of the assessors was rather vague ($\bar{x}$=2.33, Mdn=2, SD=0.972), often lacking clear suggestions for concrete next steps. Additionally, the assessors' guidance was mostly related to the competence of interest and not to other aspects such as the appearance of the artefact ($\bar{x}$=0.89, Mdn=1). […] The available data seem to provide insights into certain aspects, which relate to the productiveness of the students' implementation of the two key roles (i.e., assessor and assesse). In most cases, students' effort to provide feedback was supported by a specially designed tool, which explicated the relevant criteria for the competence of interest. Thus, the assessors were supported to systematically judge their peer's artefact against each criterion. An indication that could provide insights as to whether the students enacted productively the role of the peer assessor refers to the extent of internal consistency. In particular, while developing their own artefact, they were asked to address certain criteria, associated with the competence of interest. These criteria were negotiated beforehand with the teacher and were coded as criteria to be satisfied while producing the artefacts (as part of the teaching/learning materials and the respective activity sequence). One interesting indication that we focused on, relates to the extent to which these criteria were also employed by the students while assessing the artefacts produced by their peers. For this, we sought to identify whether the students who had addressed those criteria while developing their own artefact, also drew on these same criteria for evaluating the artefacts created by their peers. The data show that the students who had attended to certain criteria while developing their own artefact, did not necessarily attend to those criteria while assessing their peer's artefact. One possible interpretation for this apparent inconsistency is that students did not really appreciate the need to commit to certain criteria while assessing their peer's artefacts. An alternative interpretation is that they did not appreciate that coordinating the process of the development of their own artefact with the process of assessing their peer's artefacts, could serve to systematise and facilitate the latter." (D5.11, p. 33-47)

Findings related to on-the-fly interactions:

"The coding of 11 instances of classroom dialogues during the three rounds of interventions in Cyprus, revealed a total of 189 completed ESRU cycles (51% of the instances in which such a cycle started) and 182 broken ESRU cycles […] Coding the 9 instances from the 3 rounds of implementations in France (Grenoble) revealed a total of 200 completed ESRU cycles (40% of the instances in which such a cycle started) and 294 incomplete ESRU cycles. In a similar manner, coding 18 instances of dialogue from the teaching interventions in Finland, revealed that the ESRU cycle was completed 135 times, a number that corresponds to 56% of the instances in which such a cycle started. In the remaining 44% of the cases (185 instances) the cycle remained incomplete, either as ESR or ES. It is worth noticing that the percentages of complete and incomplete ESRU cycles in these countries are rather similar. […] The available data provided indications about possible reasons why certain instances of ESRU cycles were successfully completed. We were able to identify six key features, which seemed to have led to the completion of ESRU cycles" (D5.11, p. 60-61):

1) "Teacher suggests an activity in order to investigate a hypothesis that emerged during the discussion.
2) Teacher poses a question that is intended to promote students' thinking about the topic being discussed
3) Teacher poses clarification questions to help students further articulate a seemingly valid contribution they made

4) Teacher takes the opportunity to use what the students had said in order to conclude/provide feedback
5) Students expressed several ideas that allowed the teacher to ask for comparisons between these ideas
6) Teacher gets a contribution from the students and asks for further explanation with the intent to help them evaluate and adjust their reasoning " (D5.11, p. 61)

"In each of these six cases, it is evident that the teacher is seeking to promote further thinking by the student or students, which is likely to lead to richer dialogue. Sometimes this allowed the teacher to obtain more feedback from students about their reasoning. In other cases, it served to challenge students' understanding so that they could adjust the direction in which the inquiry was developing. In yet other instances, the teacher raised questions to deepen and continue the dialogue between students and student groups so that the ideas emerging from these scenarios could serve as both guidance and as a resource for other students. Similarly, we categorized our interpretations for the cases where information that came up in the discussion did not seem to be used in an explicit manner and led to incomplete ESRU cycles" (D5.11, p. 61-62):

1) "Teacher gets the answer s/he had been looking for but s/he doesn't ask for further explanation, while it would have been useful to do so
2) Teacher ignores/rejects/interrupts student's contributions to the dialogue (possible because s/he either considered them irrelevant/wrong or was looking for a different contribution)
3) Teacher offers the right answer or guides students to say the correct word instead of promoting their thinking or questioning their ideas
4) Teacher poses a vague question that students might have not been able to construe in a meaningful manner
5) Teacher poses a closed yes/no question
6) Teacher rejects an answer as non-compatible with observation/accepted theory that s/he wants them to focus on
7) Teacher does not seem to pay adequate attention to a student who tries to express a relevant/potentially interesting question, having the intention to move to another subject" (D5.11, p. 62)

"In an attempt to shed more light into the intricacies underlying the teachers' attempts to employ interactions on the fly as a formative assessment method, we focused on instances where either important information (i.e., contributions made by the students) went unnoticed during the discussion or was not used in a seemingly optimal manner. As part of the coding process, the LWGs were expected to seek and report such instances. We then undertook to synthesize the findings reported by individual LWGs with the intent to come up with a typology of such challenges. Next, we describe the different types of challenges that were reported, at the collective level. […] Challenge 1. […]The teacher is exposed to various inputs made by the students (usually as a response to his/her own probes). Some of these contributions are more likely to support the evolution of the classroom dialogue in a productive manner. At the same time, it is always possible that elaborating on (even apparently useful) contributions might end up taking the discussion away from the inquiry focus for that activity. These two features pose an important challenge for teachers, which entails two aspects: (a) how to make optimum decisions in real time as to which students' contributions to draw upon, in the sense that they are more likely to lead to productive discourse and what aspects to suppress or ignore (at least temporarily) (b) ensure a stance that is responsive/attentive to students' contributions but also allows steering the discussion in an effective manner, toward the learning goals (e.g. offer explicit guidance/feedback, deepen the discourse around the selected students' ideas through reflective questions, narrow down the scope of the

discussion etc). Making decisions, in real time, about his/her immediately next moves in the evolving dialogue, constitutes a very challenging and demanding task for the teacher. […] Challenge 2 […] There is a challenge for the teacher on how much guiding s/he should be providing in cases that s/he recognizes that students have a particular difficulty. One of teacher's options in such cases could be to let students explore and learn from their own mistakes. This could turn out to be useful for certain occasions and certain students. However, this is not always the case. It might end up as very unproductive experience for students. […] Challenge 3. […] The discourse that unfolds in the classroom is complicated by the possibility that the inputs contributed by the individual students or the teacher, could be resting on tacit assumptions, not necessarily shared by all members of that specific learning community. This might influence how the classroom discourse evolves; thus there is a challenge associated with detecting such cases which can occur in classroom discourse." (D5.11, p. 62-70).

Findings related to Structured Assessment Dialouges (SAD):
"The SAD is a short ritualised assessment involving three distinct phases: A 5-minute student-teacher dialogue, a 5-minute peer feedback session, and finally 2-3-minutes of student self-reflection. The design of the SAD is intended to allow for combining summative and formative assessment, which is highly controversial. The argument for doing this is that formative and summative assessment has been decoupled, giving privilege to summative evaluation forms in high-stakes tests with unwanted effects as a result. Since political stakeholders are not likely to abandon summative evaluation, researchers should investigate forms of assessment that both allow for learning and have a summative perspective. The structured assessment dialogue is a candidate for such a form. […] results from 26 implementations in Finland (6) and Denmark (20) [were analysed] in two ways. First, using focus group interview data, we analyse teachers' experiences with preparing, implementing, and reflecting on SAD sessions. We find that most teachers found it a useful method of assessment and reported on using it outside the scope of the study as well, although learning progressions; associated questions to ask students, and timing were seen as challenges. Second, we apply network analysis on student-teacher dialogues to produce *dialogical maps.* These maps are then grouped via cluster analysis, and groups are linked to quantitative student self-reflection measures, quantitative teacher self-reflections and contextual data. We find six different groups of dialogues each of which have different emphases. Map groups display significant differences in terms of quantitative student self-reflections, but not in terms of quantitative teacher reflections. In terms of contextual data, we find that groups of maps are linked to a high degree to the teachers that performed the dialogue." (Dolin et al., in preparation)

Work Package 6 - *Transform Results into National Contexts*

Deliverables:
D6.1 Network analysis method
D6.2 Guidelines for the function of stakeholder panels
D6.3 Reports with NSP feedback
D6.4 Illustrative examples
D6.5 Assessment Transformation Package

Objectives (from the proposal):
- To assure a national stakeholder involvement throughout the project.
- To align the research processes with the understanding and perspectives of the stakeholders for enhancing the feasibility of the results.

- To secure a meaningful communication between researchers, teachers, and stakeholders (including policy makers).
- To refine the project's assessment methods and procedures based on stakeholder feedback.
- To select the main findings and describe them in easily understandable formats.

Selected results from National Stakeholder Panel (NSP) Meetings:
"The majority of panels agreed that assessment is an important aspect of teacher education and TPD. However, in one of the partner countries it was mentioned that professional development programs for in-service teachers are not compulsory and thus there is no formal professional assessment aspect. A successful implementation of formative assessment depends on several factors, and questions on the assessment in the context of new curricula and competences are important (How do we assess competence-oriented teaching?). The implementation of formative assessment is hindered by obsolete curricula that do not provide teachers with sufficient orientation, especially related to process-oriented competences. Some teachers have the impression that they teach competence-oriented, but they do not know how to handle the assessment in relations to that; they are willing, but not educated to do so. Moreover, formative assessment requires job and life experience as well as a solid content knowledge.
What is needed in teacher education and TPD:

- Instruments, tools, guidelines, and examples of good assessment practice. However, it is not sufficient to provide teachers with diagnostic instruments - they also have to understand these instruments and the underlying principles.
- Teachers need to be convinced that they can handle competence-oriented formative assessment.
- Clear competence descriptions that could be used as a basis for formative assessment are missing.
- Regarding in service teachers, there is a need for teaching innovation projects that integrate teaching institutions (e.g. schools) and research collaborative groups.

However, it is important to note that it is not sufficient to provide teachers with materials and discuss these in short-term (e.g. one day) TPD activities. The implementation has to be accompanied in practice by long-term TPD. There is a need for alignment between the theoretical content that pre-service teachers learn in the various phases of teacher education (i.e. what we expect them to do) and the reality they find in practice (i.e. what they are able to implement). If the discrepancy is too big, the acceptance of assessment will decrease. […] The national stakeholder panels report on various ways in which summative and formative assessment are combined in practice, e.g. students' work on projects is assessed formatively during the work and is assessed summatively at the end (the summative assessment is then oral or written feedback), or several summative tests can be used during a teaching unit/a course and serve formative functions. However, it is generally agreed that consistency is important, and thus there needs to be an alignment between teaching and assessment (same criteria and visible (for students) criteria). Related to the aspect of alignment, it was mentioned in one of the panels that "that there is lack of systematic implementation of the two types of assessment. Hence, combining the two types becomes an ever more difficult task". Nevertheless, "a primary school teacher claims that the inspectors pursue the combination of formative and summative assessment via professional development programs and meetings with professionals". It was further mentioned, that at some points, and for some purposes, assessment can (and should) only be formative, and it is important to keep in mind that formative and summative assessment serve different functions. In one of the panels, it was argued that the only way to combine formative and summative assessment is by evaluating student portfolios in order to monitor students' learning progress. […] In the large

majority of the panels, they mentioned nation-wide/region-wide assessments. However, it was emphasized that most of them are usually not high-stakes for the students. Nevertheless, it was further mentioned that tests during the school year in one partner country determine students' grades for the semester, and if those grades are below average, then the final examination becomes critical and important for the students in order for them to advance to the next grade. In some partner countries, the central final examinations are high-stakes for the students. In other partner countries, it was mentioned that the roles of the tests may be increased in the future. […] In some NSP's it was emphasised that changes in assessment culture should be adapted to the relevant context – not just in terms of national context, but also educational level, subject etc.: Czech Republic NSP member: "I am a little afraid of the "quick" and universal conclu- sions" German minutes: "The panel feels that it depends on the school type as well as the grade level whether and in which form the use of formative assessment is sensible." French [minutes]: In particular, there's a similarity in the lower secondary Level assessment (summative or formal) between the students who will follow general curriculum and thus who will attend more professional specialized schools. This favours the formers. This implies a need for a change that allows a better balance in the way the evaluation of all students in the lower secondary is done. Discussion among some panels about how focusing on FA does not mean that SA hasn't got a role to play, especially since parents expect SA: Czech NSP member: "We are not in the stage where the formative assessment could be part of everyday teaching. There is still prevailing demand for grading – the teachers need them to make final certificate and parents are used to work with them too" French minutes: Also, regarding marking, there's a long history in France concerning this tradition of marking that is not easy to change. One participant mentioned that the society is quite competitive and we should educate students about it as well. German minutes: The panel speaks clearly in favour of a strict separation of formative and summative assessment to avoid a confusion of learning and achievement situations. In addition, it stresses that summative assessment cannot (and should not) be completely abolished. Finnish: Also, many parents still advocate summative testing (grades). Swiss minutes: Stronger involvement of parents: the parents want formative assessment on the one hand, but at the other hand also want to know the "worth of any artefact has" (summative assessment). There seems to be a general consensus that something needs to be changed in the assessment culture to enhance the status of FA, albeit there are different opinions about what should be changed and how this could be done. Czech minutes: The reason for the change: The school assessment doesn't support quality of students' learning with respect to understanding of content. The assessment should help student to "learn with understanding" and achieve better understanding of the content. French minutes: In general the NSP indicated that, on one hand, there's a need to engage teachers in an attitude that foster more assessments for learning. On the other hand, they indicated that the (official) educational system position is quite heterogenic regarding assessment. German minutes: The panel feels that there had been an assessment culture at schools once but it has gotten lost to a huge extent. If it would be possible to reimplement it, this would have positive influence on school development, teaching and learning. In this context, the school leaders are crucial. Swiss minutes: Large-scale assessments with innovative assessment formats could initiate more innovative teaching in the classrooms (positive teaching to the tests). So far, the existing regional-wide assessments are rather traditional. So one could also fear that more such tests kill innovative and creative teaching. Cypriote minutes: On the contrary, the majority of existing diagnostic tests are devoted to measuring students' content knowledge, without giving any emphasis to their attitudes and skills. In addition, students' assessment should be based on a wide variety of tools and methods, such as portfolio, individual and cooperative work." (D6.3, p. 3-5).

Finnish minutes: More emphasis should be put on feedback and focus more on learn- ing (what is learnt instead of what is not learnt). Generally, people should be more aware about the diversity of assessment methods and practices.

Work Package 7 - *Promote Guidelines and Results*

Deliverables:
D7.1 General roadmap for dissemination and awareness activities
D7.2 Guidelines for teachers
D7.3 Recommendations for policy makers and other stakeholders
D7.4 Organisation of conference and roundtable discussions
D7.5 National dissemination and awareness activities

Objectives (from the proposal):

- To develop guidelines for promoting actions on national and supranational levels that encourage teachers and schools to integrate formative assessment strategies in IBE
- To formulate recommendations for policy makers and other stakeholders
- To maximise the impact of ASSIST-ME through a variety of dissemination activities with user communities including policy makers and other stakeholders.
- To use media exposure to support and communicate evolving project findings.

Teacher guidelines related to interactions on-the-fly:
"Interactions on-the-fly are informal oral conversations that take place during a lesson that cannot be planned beforehand, but takes place spontaneously when the teacher recognizes appropriate opportunities to support students in advancing their learning. These can take place between the teacher and the learner, teacher and groups of learners, and teacher and the whole classroom. Interactions on-the-fly are characterised by the teacher (and sometimes learners) asking questions to elicit information, and by the subsequent use of questions (open- or close-ended) or comments (exploratory or instruction) to follow-up on the information that was gathered. Interactions on-the-fly are suitable to explore students' thinking and conceptual understanding, as well as to expose alternative conception. Interactions on-the-fly provide good opportunities for formative use of learning evidence, as the teacher can follow-up on this information in real-time and decide what is the best course of action to be taken." (D7.2, p. 6)

Teacher guidelines related to structured assessment dialogues:
"Structured classroom dialogues are ritualised conversations that take place between the teacher and one focus-student, followed by a peer-feedback conversation led by a group of students. Structured classroom dialogue is characterised by the teacher asking a set of questions to which the focus-student is expected to provide an answer, and by peer-feedback questions or comments. All of these conversations take place in front of the remaining students in the class (who are encouraged to write their reflections on a structured work-sheet). Structured classroom dialogue is suitable to explore students' thinking and conceptual understanding, as well as to expose alternative conceptions. And it also provides an opportunity for developing students' metacognition and self-regulation. Structured classroom dialogue can be used both in a formative or summative way." (D7.2, p. 6)

Teacher guidelines related to teacher written feedback:

"Teacher written-feedback is a process that enables teachers to give feedback to students by writing comments on students work (written artefacts). Teacher written-feedback provides an opportunity to gather evidence of learning from an end- product (and can reveal students' thinking and understanding, as well as alternative conceptions). Teacher written-feedback can be used formatively to provide guidance towards next-steps in learning. Teacher written-feedback can be conceptualised as written dialogue between the teacher and the student that is more spread in time. Unlike interactions on-the-fly or structured classroom dialogue, which support formative action in real time, written feedback typically supports formative action in-between lessons. Teacher written-feedback can also be used in a summative way if the comments are used to make evaluative judgements on performance." (D7.2, p. 6)

Teacher guidelines related to peer feedback:
"Peer written-feedback is a strategy that actively engages students in a process of assessing their peers' written work, of interpreting it against a set of success criteria, and of providing concrete guidance (regarding next steps in learning). Peer written-feedback provides an opportunity to gather evidence of learning from both assessors and learners, as well as providing opportunities for the development of students' metacognitive abilities and self-regulation. Students written artefacts contain evidence of their thinking and understanding. In addition, the assessor's comments and guidance also provide an insight into their own thinking and understanding. Peer written-feedback can support formative action in two distinct ways. One way is by supporting students' progression in learning through their peers's comments and guidance. The other way is by creating an opportunity for the teacher to decide on the next steps that will support both assessors and assesses, and plan activities accordingly in-between lessons." (D7.2, p. 7)

Recommendations for policy-makers and Other Stakeholders:

- "A competence-oriented, inquiry-based pedagogy is important […] An inquiry-based teaching and learning approach helps young people develop critical thinking and scientific reasoning that are important in creating citizens that can make sense of the world they live in and take informed decisions. Inquiry-based teaching and learning has proved its efficacy at both primary and secondary levels in increasing interest and attainments levels in STEM subjects, while at the same time stimulating teacher motivation. The ASSIST-ME project confirms this understanding and goes further, in defining and operationalising key competencies within STM subjects in order that students can more readily utilise and develop scientific knowledge and processes. The project points at ways to implement such a competence approach in different educational cultures and recommends adjusting educational policies to make this possible.
- Focus on formative assessment to support competence-based inquiry learning […] Formative assessment provides both the time frames and opportunities to look at how learning of competencies is developing. ASSIST-ME has collected solid evidence of the huge learning potential of formative assessment methods via their goal orientation, making the learning journey visible and explicit and via supporting teachers in identifying optimum next steps in learning. But the project has also revealed that formative assessment is not an integrated part of current STM teaching and that, for many teachers and students, it is difficult to implement in a structured form. It is therefore necessary to promote a teaching approach integrating formative assessment into the classroom culture and to frame the educational condition, resources and the curriculum to make it happen.
- Reduce the emphasis on summative assessment to give room to formative assessment […] ASSIST-ME found that the summative assessment load needs to be reduced to allow

teachers time to focus more on formative aspects and to highlight and emphasise those aspects of learning that we value within the STM community. Curricular material, textbooks and resources need to include specific and detailed reference to the formative potential and use in the classroom so that both teachers and students are focused on how assessment can support learning. It is recommended to develop national assessment policies that recognise the different roles and potential involved in the interactions between formative and summative assessment and that makes it possible to realise the full potential of formative assessment processes.

- Develop new examination forms able to capture STM competencies […] The ASSIST-ME implementations have made it clear that there is a big gap, in many countries, between the examinations at the end of a course and the learning processes in the course. While the teaching-learning processes are aiming at developing the learners' STM competences, the examinations often fail to assess these properly. To bridge this gap, the summative assessments should be more in alignment with the formative processes in everyday teaching and they should be designed to assess the STM competences in a valid and reliable way. We know from the project that classroom practice is heavily influenced by a backwash from summative examinations and so the development of examination forms that assess a broader range of STM competences would also have a positive impact on the teaching of these competencies. It is necessary to develop new examination forms able to capture the central STM competencies and being aligned with the formative approaches in the classroom.
- Teachers need support in implementing and enacting classroom assessment of STM competencies […] ASSIST-ME has developed formative assessment tools able to support teachers in defining and articulating appropriate feedback comments for students, thereby strengthening the assessment literacy of both parties. However, assessment tools alone are insufficient as teachers need to adapt such tools to their educational contexts and this requires support from peers and educators in translating tools for specific use. The ASSIST-ME model may provide an effective format for these programs: three meetings per year, involving a team of teachers and researchers or teacher educators, in designing and testing new teaching units in an iterative process dedicated to the on-going improvement of the inquiry activities and their assessment tools. ASSIST-ME has identified a strong need for professional development programs (pre-service, induction and in-service) that support teacher understanding of formative assessment and inquiry-based teaching and learning and facilitate the implementation and enactment of formative assessment processes in STM classrooms at both primary and secondary level." (D7.3, p. 2-4)