

Report from the FP7 project:

Assess Inquiry in Science, Technology and Mathematics Education



ASSISTME

National reports of partner countries reviewing research on formative and summative assessment in their countries

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

Delivery date	31.07.2013
Deliverable number	D 2.3
Lead participant	Leibniz-Institute for Science and Mathematics Education (IPN), Kiel, Germany
Contact person	Silke Rönnebeck, roennebeck@ipn.uni-kiel.de
Dissemination level	PU

Table of Contents

SUMMARY	4
1. INTRODUCTION	5
2. NATIONAL COUNTRY REPORTS	9
2.1 Cyprus: Assessing Inquiry in Science, Technology and Mathematics Education – A report on research in Cyprus	10
2.1.1 Foreword	10
2.1.2 Why research on formative/summative assessment has not received adequate attention in Cyprus?	10
2.1.3 Emphases of research on assessment, more broadly, in the context of Cyprus	11
References	12
2.2 Czech Republic: Report on Czech Republic formative and summative assessment results	13
2.2.1 Foreword	13
2.2.2 Formative and summative assessment research results in the Czech Republic.....	13
2.2.3 Formative and summative assessment in inquiry-oriented education in the Czech Republic (summary of answers to WP2 questions)	17
2.2.4 Overview of publication dealing with assessment in Czech Republic schools	18
References	20
2.3 Denmark: Review and syntheses of Danish research results on assessment – Danish LWG	23
2.3.1 Answers to the questions	23
References	32
2.4 Finland: Report from Finland concerning formative and summative assessment with respect to IBE and STEM	34
2.4.1 Foreword	34
2.4.2 Answers to the questions	34
References	39
2.5 France: Contribution of ICAR – CNRS, France	41
2.5.1 Answers so the questions	41
2.6 Germany: A synthesis of the research on assessment of IBE	51
2.6.1 Foreword	51
2.6.2 Answers to the questions	51
References	63
2.7 Switzerland: Synthesis of Swiss Research Results on Formative Assessment	66
2.7.1 Summary	66
2.7.2 Answers to the questions	66
References	74

2.8 Report of Assessment of STM in England	77
2.8.1 Introduction.....	77
2.8.2 Answers to the questions	77

ANNEX - QUESTIONS TO GUIDE NATIONAL COLLECTION OF RESEARCH (D2.1)	85
--	-----------

Summary

The EU-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.

In order to get an overview of the state-of-the-art of formative and summative assessment in inquiry-based education (IBE) in STM in the ASSIST-ME partner countries, the countries were asked to collect national research related to this issue. The collection was guided by 10 questions provided to the countries by work package 2 as deliverable D2.1. This report presents the reports produced by the countries and summarizes some major findings.

The country reports indicate that the situations in the countries differ significantly both with respect to the implementation of IBE and formative and summative assessment, respectively. Nevertheless, almost all countries state that there has not been much research concerning formative assessment (or assessment in general) of IBE and that summative assessment plays the dominant role in student assessment in STM. However, approaches to formative assessment exist in several countries and have been investigated in small to medium scale studies.

1. Introduction

The aim of this report is to supplement the international literature review on the current state-of-the-art of formative and summative assessment of inquiry-based education (IBE) in science, technology, and mathematics (STM) education. The success of any approach to change the assessment practice within a country will, at least in part, depend on country specific factors like e.g. the assessment tradition, characteristics of the educational system, or experiences with different forms of assessment. Consequently, these factors are very important for the future work in ASSIST-ME.

The individual country reports summarized in this report thus pay special emphasis to the situations found in the ASSIST-ME partner countries. The countries were asked to collect recent research that had been conducted in their countries on formative and summative assessment in STM. To facilitate this work, they were provided with guidelines (see ASSIST-ME deliverable D2.1) that formulated ten questions to guide the review process. These questions are summarized in the Annex. The main section of the report consists of the individual country reports in alphabetical order¹. In this introduction, some interesting results from the reports should be exemplified². For details, please refer to the respective country report.

The most interesting and striking similarity of the reports is that all of the countries state that there has been little to almost no national research on formative assessment of IBE (or assessment in general) in their countries. Consequently, they all found it (more or less) difficult to give research-based answers to the questions formulated in the guidelines. Countries solved this problem by either concentrating on those areas where there is research in their countries or by providing hypotheses concerning those questions where research results are missing. Some countries explicitly tried to give reasons for the lack of research in this area like e.g. Cyprus or the Czech Republic. In Cyprus, for instance, a possible reason was seen in the fact that educational policy and teaching practice do not prioritize evidence-based research in their decisions. As a consequence, the potential of assessment data to inform policy and practice (and support learning) is often ignored. The centralization of the educational system was regarded as another possible cause. Teachers often lack the motivation to improve the quality of their teaching since such efforts are not rewarded by the system (teachers usually get permanent positions and their appointment and salary are independent of their qualifications and the quality of their teaching).

With respect to the role that summative and formative assessment plays in and for the teaching and learning of STM, in almost all countries summative assessment is considered to be predominant compared to formative assessment. In some countries like Switzerland or Germany, long traditions of summative assessment and grading exist. The same is true for Finland where students have to complete up to 50 tests/year and

¹ Principally, the reports are presented as they were submitted by the countries. In some cases, however, some clarifications were necessary. They are attached to the main report under the heading 'additional information by the authors'.

² Results from England are not included in this introduction due to the late submission of the national report.

the test results often provide the major source for grading. The character of the tests differs (e.g. from nationwide to state- or even school-/teacher-based tests) in relation to the centralisation of the educational system and e.g. the school autonomy. Within the last decade, however, even in strongly decentralised systems like e.g. Switzerland and Germany, tendencies have been observed to establish a nationwide comparability of assessment tasks and results. As a consequence of the dissatisfying results of German students in TIMSS and PISA, the educational system in Germany, for instance, shifted from an input- to an output orientation. Nationwide educational standards were implemented and their attainment is monitored in regular intervals by national large-scale assessments. However, similar to Switzerland, the purpose of these large-scale assessments is to survey the system and not the individual student.

With respect to formative assessment, results from France show that the majority of teachers and students favour formative assessment or at least consider formative and summative assessment equally important. In Switzerland, mandatory guidelines exist in many cantons that explicitly mention formative assessment; however, no systematic surveys of formative assessment practice exist (which seems to be characteristic for the situation in many of the partner countries). Moreover, a common understanding between teachers and politicians that assessment and lesson planning should be oriented at educational goals and competences still seems to be missing.

Structured formative assessment seems not to exist in the ASSIST-ME partner countries. A recent study in Denmark, however, found that when teachers assess their students, they have “an outspoken focus on learning and learning potential” and that most teachers assess “continuously and after the individual activity” – the most common forms of assessment are whole class conversations and written tests. A similar hypothesis in Finland assumes that opportunities for formative assessment exist in daily teaching practice. Teachers might use especially short-term, informal formative assessment in teacher-student interactions. In Switzerland, student and teacher attitudes towards different assessment methods have been investigated. A high acceptance by teachers for oral feedback instead of grades, learning reports on progress, and student self-assessment could be observed. The acceptance among parents and students was also comparably high.

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). In Germany, a study is

underway that aims at investigating the relationships between teacher characteristics, their instructional and assessment practices, and their use of formative assessment.

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. This training might address different aspects. On the one hand, teachers need support to increase their 'assessment literacy'. Research e.g. from Germany shows that a high diagnostic competence of the teacher positively influences his or her formative assessment practice. Moreover, teachers need support to change their beliefs about assessment (see above). 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. 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. 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.

In all countries, there is not much research or information about the interaction between formative and summative assessment available. Grades are regarded by some countries as a form of summative assessment that also has a potential for formative assessment purposes. A study in Germany comparing classroom based assessments and standard-based tests found that thematically focused assessments – as needed for formative assessment – led to additional and specific information that could not be provided by summative assessments.

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. One 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'). A study in Germany 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 Germa-

ny, 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 there are lots of instruments available but very little research-based knowledge on how they are used. 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. With respect to IBE, several small to medium scale studies in Germany focused on the (summative) assessment of experimentation competence in science education. In contrast, a Danish formative assessment instrument aimed at supporting students in performing inquiry processes in physics. It was shown to increase the motivation, especially of girls, dramatically.

In most of the countries, research results on assessment either are not available or not used by anyone outside the research community. An influence of large-scale assessments like TIMSS and PISA could be observed in Switzerland and Germany. In both countries the results provoked discussions. While in Switzerland the assessment practice nevertheless remained unchanged, the discussions in Germany led to significant changes at all levels of the educational system.

2. National Country Reports

In the following, the national reports from the ASSIST-ME partner countries (as we received them) are presented (in alphabetical order). Table 1 contains a list of the authors.

Country	Authors
Cyprus	<i>Elena Siakidou, Nicos Papadouris, Constantinos P.Constantinou, & Michalis Livitzis</i>
Czech Republic	<i>Lukáš Rokos & Iva Žlábková</i>
Denmark	<i>Jan Alexis Nielsen</i>
Finland	<i>Pasi Nieminen & Jouni Viiri</i>
France	<i>Florence Le Hebel , Sylvie Coppé, Suzane El Hage, Pascale Montpied, Andrée Tiberghien, & Zeynab Badreddine</i>
Germany	<i>Sascha Bernholt, Silke Rönnebeck, Mathias Ropohl, Olaf Köller, & Ilka Parchmann</i>
Switzerland	<i>Regula Grob, Monika Holmeier, & Peter Labudde</i>
England	<i>Christine Harrison</i>

2.1 Cyprus: Assessing Inquiry in Science, Technology and Mathematics Education – A report on research in Cyprus

Elena Siakidou, Nicos Papadouris³, Constantinos P.Constantinou, & Michalis Livitzis

2.1.1 Foreword

We would like to state, from the outset, that research on assessment in education (either for formative or summative purposes) has been very scarce in Cyprus. Thus, there are not any original, published research papers reporting results, situated in the context of Cyprus that could be relevant to the questions provided to guide the process of structuring this report. We have decided to organize this report in two sections. The first, reviews the main reasons why formative/summative assessment has received scant attention as a research field. The second, outlines research areas, falling within the field of assessment more broadly, that have received some attention and briefly reviews some key findings that have emerged.

2.1.2 Why research on formative/summative assessment has not received adequate attention in Cyprus?

We believe that a main reason why formative/summative assessment, as a research domain, has not received considerable attention, in the context of Cyprus, is that implementing evidence-based policy and seeking evidence-based decisions on how to enhance the quality of teaching and learning are not included among the priorities of either the educational policy or the conventional teaching practice. This could be partly attributed to particular characteristics of the educational system, which are detailed next. One of these characteristics is that the educational system does not differentiate between schools on the basis of academic criteria, which is common place in other countries. Thus, there is not any measure that could be used to compare schools, teachers or students across the educational system (especially in the public school system, which is the most common choice for students). This tends to somehow weaken the motive for school units (or individual teachers, for that matter) to demonstrate enhanced quality. Even though students are tested at various occasions throughout the academic year, this is typically done with the sole purpose to assign them a score and determine whether they have passed or failed a given course, rather than to collect evidence that could be used to adapt teaching so as to better meet their needs. Thus, the role of assessment tends to be restricted to obtaining a ranking for the students who have taken a specific test, and (often) to comparing their score with a standard criterion, which determines success or failure on the test/course. The idea that data from assessment could be also used in useful ways other than merely ranking and/or comparing students is severely underdeveloped. This is evidenced by the university entrance examinations that are organized in Cyprus each year. This yields a large corpus of data on students' achievement, whose facility to provide useful feedback for both educational policy and teaching practice is typically ignored. The data are solely used to rank students with the intent to determine placement in the various academic

³ Corresponding author (e-mail: npapa@ucy.ac.cy)

programs whereas their potential to provide valuable insights that could be used to inform both educational policy and the teaching practice is typically ignored.

Another characteristic is that the educational system is highly centralized. Teachers in public schools, nationwide, are essentially appointed by a state authority, namely the *Educational Service Commission*. Appointment is done in a manner that gives priority to how early one has obtained her degree, rather than on criteria relevant to qualifications that are actually linked to competence in teaching specific subject matters. In particular, upon completion of a degree that qualifies for entering the profession of a teacher (either at the primary or the secondary level) one applies to be registered in a list of candidate teachers. The primary criterion that determines the position of a candidate teacher in this list is the year when she obtained her degree. When appointing new teachers, the *Educational Service Commission* is obliged to pick those who are in the highest position in this list. Another relevant characteristic is that getting an appointment, is essentially equivalent to a permanent job. The salary of each teacher is solely determined by how many years she has been working for and is totally dissociated from specific qualifications or performance as a teacher in the class. This implies an environment without sufficient motivation for teachers to demonstrate “good teaching practice” so as to retain their position or pursue a higher salary. Another characteristic, which also stems from the educational system being highly centralized, is that students do not get to choose the school in which they would like to study (except for the private schools, which implement selection procedures based on entrance examinations). Each student is supposed to study in a certain school according to the address of her residence. These characteristics could, at least partly, account for why formative/summative assessment has not received substantial or explicit attention as a research topic. Also, they explain why the value of formative assessment, as a powerful resource for enhancing the quality of teaching, is not usually assumed and taken advantage of in conventional teaching practice. Indeed, while the collection of information on students’ learning outcomes is common practice in conventional teaching this information is not typically used to inform teaching.

2.1.3 Emphases of research on assessment, more broadly, in the context of Cyprus

Even though research on assessment of students’ learning outcomes on various subject matters (including Natural Sciences, Mathematics, Design and Technology) is practically inexistent, it is the case that there has been some published research on topics associated with assessment, more broadly. One of these topics involves the evaluation of the effectiveness of the school unit and the teachers. The following are some of the main findings of this research: (a) one of the most important factors that influence the effectiveness of the school unit involves the establishment of a mechanism for its internal evaluation (Creemers, Kyriakides, Demetriou, & Antoniou, 2007); (b) a school unit is more effective when it organizes and implements activities targeted at the improvement of its teaching practice and the corresponding learning outcomes (Creemers & Kyriakides, 2006). Another research area that has received some attention relates to the investigation of teachers’ views about assessment. Studies in this area suggest that the dominant view is that assessment provides a means for improv-

ing both the quality of teaching and learning (Kyriakides, 1996; Michaelides, 2009). Despite this, other studies have suggested that teachers tend to assume more traditional assessment approaches, like written tests, in their own teaching practice. Thus, while they do seem to appreciate assessment as an integral component of teaching and a powerful means of enhancing the quality of teaching and learning, they exhibit an inclination towards traditional assessment approaches that yield overall scores (rather than detailed information on students reasoning), which could be more meaningfully construed by parents. Finally, another area that has received some attention includes peer-assessment skills and the extent to which it could improve understanding of the content. (Tsivitanidou, Zacharia & Hovardas, 2011).

Additional information from the authors

In the Cyprus educational system there is a subject titled "design and technology". This is a compulsory course for students in the elementary school and also for the three lower grades of secondary education (students aged 6-15). Also it is an elective course for students in the upper grades of secondary education (students aged 15-18).

References

- Kyriakides, L. (1996): The perceptions of Cypriots about the assessment of students' achievement (in Greek). *Pedagogical Review*, 24, 261-283.
- Creemers, B.P.M. & Kyriakides, L. (2006): A critical analysis of the current approaches to modelling educational effectiveness: the importance of establishing a dynamic model. *School Effectiveness and School Improvement*, 17(3), 347-366.
- Creemers, B.P.M., Kyriakides, L., Demetriou, D., & Antoniou, P. (2007): *A critical analysis of the dynamic model of educational effectiveness based on a synthesis of studies investigating the impact of school factors on student achievement gains. Paper presented at the ICSEI 2007 Conference*. Portoroz (Slovenia), January 2007.
- Michaelides, P., M. (2009, July): *A survey of teachers' assessment practices and conceptions on a Cypriot sample*. Poster presented at the 74th Annual Meeting of the Psychometric Society, Cambridge, UK.
- Tsivitanidou, O.E, Zacharia C. Z & Hovardas, T. (2011): Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21, 506-519.

2.2 Czech Republic: Report on Czech Republic formative and summative assessment results

Lukáš Rokos⁴ & Iva Žlábková

2.2.1 Foreword

No research focusing on formative and summative assessment was conducted in the Czech Republic in the past fifteen years. Informations about the state of assessment are available in the OECD report regarding education in the Czech Republic as well as in research that involves the education process and communication in education. For that reason we include research results as they offer relevant information about the state of assessment in schools in the first part of our report. On the basis of these results we reply in the second part of the present report to the questions raised by the WP2. Finally, in the third part, we present an overview of publications that deal with assessment in the Czech Republic.

2.2.2 Formative and summative assessment research results in the Czech Republic

The report of the Organization for Economic Cooperation and Development (OECD) indicates that the students', teachers', schools' and educational systems' assessments in the Czech republic are not mutually consistent and interconnected and that there is a need to lay down an integrated assessment framework conception and to determine the way of how to arrive at interconnectedness among the individual constituents. Due to the fact that the Czech educational system is significantly decentralized and dependent on evaluative abilities of various agents it is of great import that the development of such an integrated assessment system reflects various needs of school management organs (on a regional and a local level), of school principals and educators (comp. Santiago et al., 2012).

In the conclusion of the OECD report that focused on the process of education it is stated that the teachers enjoy in the process of assessment a lot of autonomy. However, in the teaching organization and in the assessment process dominates a traditional approach. The assessment is most of the time summative. Through it the students receive a feedback on their performance at the time when they already cannot improve their results. Formative assessment that supports the students' learning process is not systematically utilized in Czech schools. Among the negative phenomena belongs also the fact that the assessments in schools and in classrooms are not consistent enough because the responsibility for establishing the criteria for student assessment lies with the school and that is the reason why the rules and the procedures in individual schools differ from each other and are very general. The individual teachers' criteria for grading differ as well. At the same time Czech teachers are not used to specify in detail their criteria and to communicate them to the pupils in advance. Assessment whose purpose is to assist the students in their learning is not systematically applied in the Czech Republic. Evaluative procedures insufficiently emphasize providing feedback to students and the development of learning interactions between the teacher and the

⁴ Corresponding author (e-mail: rokosl02@pf.jcu.cz)

student, interaction that facilitate learning. Feedback is frequently understood as a “summative (output) assessment that is performed more frequently”. Among the main recommendations in the area of assessment in education belong:

1. reinforce teachers’ competencies to evaluate students (formative and summative),
2. conscious application of formative assessment that supply the students with information about how to improve their learning,
3. work out national methodological directives for assessment in accord with the educational objectives in order to reduce differences in assessment and grading of students by individual teachers (compare Santiago et al., 2012).

The state of assessment in education is confirmed also by the results of multiple case research studies that were conducted in the years 2007-2009 in five elementary schools. The results that focused on assessment in the schools show that the main assessment agent is the teacher, that the predominant form of assessment is grading, and that the main source for results assessment is a written test. Verbal assessment as a substitute for grades did not occur in the schools; verbal comments appeared in a few cases when they accompanied the grade but these were more appreciative statements expressing the teacher’s emotions than clarifications of the student’s performance or mistake identifications (Dvořák et al., 2010, pp. 242-248).

The teachers used various grading systems. Some teachers emphasized “objective” comparisons of students among themselves; others evaluated the students according to given criteria. The teachers had their own image of an ideal achievement and they usually had a clear idea about criteria assessment. However, in the studied classes there were no records of introducing the students to the assessment criteria at the beginning of the classes. Occasionally, the assessment criteria were expressed explicitly as when the teacher justified a specific grade during oral examination (Dvořák et al., 2010, pp. 245-246).

Teacher’s assessment predominated in the class. Self-assessment and student reciprocal assessment took place in the auto-correction of the mistakes. The students by themselves - or one by the other - identified their mistakes and recorded their results. On some occasions a student was asked to assess his or her own performance and this self-assessment was confronted with the teacher’s. In the framework of the teaching hours appeared praises as a type of performance appreciation (Dvořák et al., 2010, pp. 245-246).

K. Šedřová, R. Švařiček, and Z. Šalamounová (2012) in their publication *Komunikace ve školní třídě* (Communication in school classroom) pursue the issue of assessment with providing feedback to the student. They present data from their research demonstrating that within the context of feedback to the students during dialog communication the most frequently employed mechanism is the - so called - *unsaid evaluation* when the student’s answer is not explicitly evaluated and the students infer that their response was correct from teacher’s passing on to the next question. Assessment is unpronounced; it is hidden in the shadow of the new encouragement. It steps out of the

shadow only at a moment when the student's reply is incorrect and the teacher needs to clarify that. Instructive evaluative comments are, nevertheless, quite rare even on such occasions. In case the teacher asked a closed question, he or she most frequently encouraged the student to give a new answer or mechanically prompted the student. In case the teacher asked an open question, he or she gave up right away on any assessment and accepted all student's answers as valid without further elaboration. In such a way the teacher renounced the possibility of using the formative potential of students' replies (Šedřová, Švařiček, and Šalamounová, 2012, pp. 132-133).

In addition, natural science video study by TIMSS 1999 also dealt with an assessment research and its results were published in the year 2006. This research analyzed and evaluated video recordings of natural science classes in the eighth grade. The analysis of video clips focused on class organization, on the way of subject presentation, and on the option of students joining in and participating in the class. Five countries participated in the nature science part of the study: Australia, Czech Republic, Japan, Netherlands, and U.S.A. For the Czech Republic 88 class-hour video recordings were included in the analyses; among them were geography (28 hours), biology (18 hours), physics (16 hours), chemistry (17 hours), and 20 other hours (interdisciplinary classes, technology and society, character of natural sciences, natural sciences and mathematics). The results show that in the Czech Republic 9% of the teaching time is devoted to examinations. In comparison to other countries the Czechs spent most time in repetitions of the studied material and in examinations. In the analyses of class hours were included observations as to how often were the students evaluated publicly, how often they were graded, or how often they had to work under others' scrutiny (e.g., at the blackboard). Public grading appeared in 19% of classes, public assessment in 37% classes and work under others' scrutiny in 75% of classes. Czech Republic was significantly higher in this area than the other countries. In the Czech Republic, in the natural sciences education, the emphasis is put on the acquisition of knowledge in the form of facts, definitions, and formulas. Lesser attention is being paid to the search for related contexts. In comparison to other countries Czech students encounter a great number of scientific findings. Only about 14% of the whole education time is devoted to individual or group work. In the other participating countries this type of work took 42-48% of all education time. Furthermore, another aspect that was examined was how much time is dedicated to longer term projects. In the Czech schools longer term project almost did not appear (comp. Mandřková and Palečková, 2007, pp. 238-248).

On the basis of these analyses, a characteristic profile of teaching was generated for each one of the countries. What characterized the Czech Republic was that in the natural science classes the teacher works with all the students in the class together and concentrates most of all on whether the content of the passed on knowledge is correct. A typical class consists of repetitions, examinations a passing on of prescribed elements of knowledge to the students and it has relatively little time invested in independent practical activities of the students. A salient characteristic of the Czech classes are the repetitions and oral examinations. The hour load is demanding and theoretical and involves mostly acquisitions of facts and definitions instead of examining any interdependencies among them (comp. Mandřková and Palečková, 2007, pp. 248-249).

Based on the video studies of TIMSS 1999, *Centrum pedagogického výzkumu PedF MU* (Center for Educational Research at Education Faculty, Masaryk University) in Brno conducted video studies of physics, geography, English language, and Physical Education classes. From the observations of the results in physics classes video studies that were conducted in the academic year 2004-2005 in 19 classes (7th and 8th grade) with 13 teachers in 12 elementary schools in Brno, conclusions similar to those from the video studies of TIMSS 1999 were arrived at. On the average 12.2% of the whole class-time was devoted to examinations, quizzes or to homework checking. Into this category belongs also oral examination at the blackboard that is for us – in contrast to other countries – natural part of our school life. In some classes a time was devoted to writing quizzes; the most time was devoted to checking how correct are the tasks' solutions that were worked out either as a part of an home assignment or in an individual or group work in the class (comp. Janík et al., p. 42, p. 46).

Part of the research was to investigate how much time is devoted to the –so called- recapitulation phase. The research authors conceive of recapitulation as of a repetition that does not focus on the content and, instead, examines the undergone work or learning process. The student should be conscious of what he or she had to do to solve the given problem correctly and this should enable him or her to transfer the employed procedures to new situations. The teachers are expected to support the students' learning process. This recapitulation phase was missing in most classes in the research sample; it formed on the average only 2.2% of the whole class time. In case it appeared, the teacher implemented it in the form of an explanation or it took place in a discussion with the class.

The video studies of geography lessons were conducted in the years 2005-2007 in the cooperation with the Geography department and the *Centrum pedagogického výzkumu PedF MU* in Brno. The objects of analysis were 50 teaching hours of geography whose topic was "The natural conditions of the Czech Republic". The data were collected from six teachers in 8th and 9th grade classes in elementary schools in Brno and its close environs. The most time was spent on the examination/quiz/home assignment check phase (9 minutes 56 seconds). During this phase most of the time was spent on oral examination of student/-s at the blackboard. Written examination was conducted at the end of a thematic unit. The solution correctness of the assigned home work was checked at the beginning of the class in the form of a discussion with the class. Recapitulation phase was in most classes missing. If it appeared, it was performed by the teacher in the form of a lecture/explanation/instruction (comp. Hübelová et al., 2008, p. 54, pp. 63-64, p. 67).

From the discussed research findings it can be concluded that the option of formative assessment is not fully utilized. Despite that fact that wider studies that would deal with assessment in Czech schools have not been published as yet, it is possible to assert on the basis of the mentioned sources that Czech teachers rely in most cases on summative assessment. Up to now they insufficiently utilize formative assessment options and do not provide their students with feedback that would support the development of a more effective learning process. Without a doubt teachers in some innovative

and “alternative” schools use formative assessment, none-the-less, it is not a country-wide phenomenon.

2.2.3 Formative and summative assessment in inquiry-oriented education in the Czech Republic (summary of answers to WP2 questions)

Inquiry-based education is, as of today, not a regular part of natural science subjects learning in the Czech Republic. Since no studies that would focus on the exploration of formative and summative assessment in nature science subjects were conducted, we are unable to give exact answers to these questions. On the basis of the assessment publications screening, and on the basis of the above discussed research results, we arrived at the following conclusions.

From the studies that deal with the process of assessment in education can be concluded that summative assessment in education is predominant and large time space is devoted to it. Teachers systematically fail to utilize feedback as an instrument for further student’s learning process regulation.

The results of the studies are published in educational professional journals and are available to the wide education public. However, it cannot be ascertained to what degree are read and applied by teachers. Teacher specialized journals tend to publish partial experiences of teachers from their practice.

In order to support the introduction of formative assessment into the education in the Czech Republic it is necessary to work out the issue of formative assessment on both theoretical and methodological level. There are an insufficient number of publications available to teachers; those are publications that should offer them unified theoretical and methodological basis for the application of formative and summative assessment. Insufficient theoretical and methodological basis is a great hurdle preventing the introduction of formative assessment both in inquiry-based learning as well as in the usage of other learning strategies.

In the publications that are available to teachers the issue of formative and summative assessment is presented only on a general level. Some publications include concrete examples of implications resulting from formative and summative assessment in education practice; these are, however, in most cases only illustrative examples. Moreover, almost none of these examples are specifically oriented on the subjects with natural science character or on the options of employing them in inquiry-based learning.

Summative assessment can be used as a support for formative assessment in case it is either a means to communicate to the student feedback information as regards the undergone learning process, or in the case the student’s performance is analyzed, or if he or she is shown ways of how to overcome the failures.

There are no elaborated methodological materials for inquiry-based activity assessment in the Czech Republic. In consequence, it can be assumed that teachers use during their inquiry-based activity summative assessments only at the end when the students’ overall achievements are being assessed. Usage of formative assessment that would facilitate the assessment of individual steps during the students’ activities

could improve the adoption of new methodological procedures based on inquiry-oriented learning.

There is no data basis of instruments and protocols either for formative or for summative assessment in the Czech Republic.

2.2.4 Overview of publication dealing with assessment in Czech Republic schools

There are no books that would specifically treat the topic of formative and summative assessment in STM in the Czech Republic. There are available four books dealing in complex ways with the issues of assessment in education. Further, this issue appears in teachers' handbooks that deal with the issues of didactics, psycho-educational diagnostics, and school psychology.

A crucial publication, *Hodnocení v současné škole* (Slavík, 1999, Assessment in the present school), brings a complex look at the issue of school assessment. It characterizes the assessment process in school, the relationship of assessment to the concepts of learning and to objectives of teaching; it treats in detail assessment criteria and the use of norms during assessment, the issue of assessment objectivity, and it supplies concrete suggestions for teacher's work. The whole publication is intertwined with the topic of formative assessment. The author presents ways to work with the given assessment system that could be utilized for formative purposes on the path to the student's autonomous assessment.

In the year 2005 came out the publication *Hodnocení žáků* (Student assessment) of Zdeněk Kolář and Renata Šikulová that deals with the outlining of student assessment with the focus on its specifics and it delineates school assessment functions. It treats in detail the education process phases with its forms of assessment, objectivity and subjectivity in school assessment, with the meaning of positive and negative assessment in school, with the roles of educational process participants, and with students' self-assessments. In the year 2009 a new edition of the book came out and was enlarged by one chapter that deals with the assessment of student competencies. It gives an integrated overview of the assessment process in education (Kolář and Šikulová, 2009).

In addition, in the year 2008 came out a handbook for teachers, *Školní hodnocení žáků a studentů se zaměřením na slovní hodnocení*, by Hana Košťálová, Šárka Miková, and Jiřina Stang. It discusses the system of -so called- assessment for learning and its uses in the process of authentic and manageable learning. Assessment for learning is supposed to help students understand their own learning process and to guide the student to the improvement of both the process and the results of learning. It emphasizes the use of a descriptive language instead of the, currently frequent, judging language. It deals in detail it deals with students' self-assessments, with criteria and indicators of assessment, and with the place of student's portfolio in the process of assessment. It offers examples of data collection methods that gather data about student learning and that serve as a basis for ongoing learning correction. In its conclusion it discusses the

use of student-teacher consultations within the framework of the assessment process (Hana Košťálová, Šárka Miková, and Jiřina Stang, 2008).

In the year 2011 came out a publication *Systém hodnocení a sebehodnocení žáků* (System assessment and self-assessment of students) from Jana Kratochvílová. The author describes in her publication a complex system of conceptualizing the developing assessment and self-assessment of students in elementary schools. The book bases itself on the long-term experiences from Czech and European (Brussels) schools and from research insights. It deals with outlining the process of assessment as it relates to the conception of education, functions and types of assessment, forms of assessments, evaluative language, criteria of assessment, competencies assessment, and with students' self-assessments. In the last chapter the author proposes student-teacher interconnected system of assessments involving assessment of the student by the teacher as well as the student's self-assessment. It is done in such a way that both types of assessment complement each other and offer a complex view of the student's learning process and his or her results (Kratochvílová, 2011).

A handbook for teachers that reacts to an actual demand for developing and for assessment of student competencies is the publication *Nápady pro rozvoj a hodnocení klíčových kompetencí žáků* (Ideas for development and assessment of key competencies) from Barbara Hansen Čechová from 2009. In this publication - in continuation to the development of student key competencies - the author describes tools for the development of self-assessment (Hansen Čechová, 2009).

Furthermore, various chapters dealing with forms of assessment appear in didactics oriented publication of Czech authors such as in *Obecná didaktika* (General didactics) by J. Skálová (2007, pp. 209-216), in *Školní didaktika* (School didactics) by Z. Kalous, O. Obst, et al. (2001, pp. 403-414), in *Pedagogika pro učitele* (Pedagogy for teachers) by A. Vališová, H. Kasíková, et al. (2011, pp. 249-266), in *Analýza vyučování* (Analysis of teaching) by Z. Kolář and A. Vališová (2009), and in *Úvod do didaktiky základní školy* (Introduction to elementary school didactics) by O. Šimoník (2005, pp. 118-125). These publications deal most of the time with instrumental assessment of students. Assessment is considered to be an instrument that enables the teacher to externally control activities. The particular student's assessment motivates, directs, and organizes his or her learning activities.

M. Chrátka (1998) in his *Didaktické testy* (Didactic tests) deals with assessment tools and their usage. The issues of achievement tests are discussed in a wider context of students' and school diagnostics in publications dealing with psycho-educational diagnostics such as in *Diagnostika* (Hrabal st. and ml., 2002, Diagnostics).

Instructions for student assessments in the class together with concrete examples of their applications are presented in publications intended for teachers who are interested in developing their teaching skills such as in *Pedagogika ve škole* (Starý et al., 2008, p. 47-69, Pedagogy in school) and in *Učitelé učitelů* (Starý et al., 2008, pp. 65-87, Teachers' teachers). The authors of these publications deal with criteria for evaluating and they present examples of their usage in the work of a teacher.

V. Kosíková (2011) in her publication *Psychologie ve vzdělávání a její psycho-didaktické aspekty* (Psychology in education and its psycho-diagnostic aspects) discusses psycho-didactic context in the process of school assessment. She focuses on the conceptualization of assessment, mutual connectedness among educational and evaluative objectives, formative function of assessment, and on work with mistakes in the evaluative process.

Czech and foreign experiences with student achievement measurements were presented in the publication *Kvalita cestou kvalifikace* (Quality through qualification) by Seebauerová, Helus, and Kolidias, (2000). This publication elaborated on the issue of achievement measurement, classification and verbal assessment, construction of test tasks, usage of portfolio, and the psychological aspects of performance assessment.

Among the foreign publications that were translated into the Czech language and that deal with the issue of assessment is the following: *K problému známkování ve škole* (As regards grading in school) from a German author, J.W. Ziegenspeck (2002). The book describes general issues in assessment and the results of foreign research. And then there are chapters about assessment in handbooks for teachers in publications such as: *Moderní vyučování* (Teaching Today) by Petty (2008), *Od vzdělávacího programu k vyučovací hodině* (Teaching as decision making) from Pasch et al. (2005), *Učíme děti myslet a učit se* (Teaching children to learn) by Fischer (2011), *Klíčové dovednosti učitele* (Essential Teaching Skills) by Kyriacou (2012), and, finally, *Strategie řízení třídy* (Classroom Management Strategies) by Cangelosi (2009). These publications came out in a number of editions.

Additional information from the authors

Few of our more-case studies were not executed only in one subject. The TIMSS studies are exceptions because they were executed in concrete subjects (Geography, Biology, Physics, Chemistry and other fields – one of them was Technology and the second one was Mathematics). Continuing studies were oriented in one subject too – these studies were performed by Masaryk University in Brno. A lot of our literature sources are written in common way so they are not directly connected to assessment in concrete subject. For example there are examples from Art in Slavík (1999), but other authors don't use examples only from one field.

We were oriented on inquiry in science field, because there is the most obvious influence of this approach in Czech conditions. Furthermore we can find components of inquiry in mathematics too. Technology is not taught in Czech educational system as self-contained subject.

References

- Cangelosi, J. S. (2009): *Strategie řízení třídy* [Classroom Management Strategies]. Praha: Portál, 5. vyd.
- Dvořák, D., Starý, K., Urbánek, P., Chvál, M. & Walterová, E. (2010): *Česká základní škola. Vícepřípadová studie* [Czech elementary school. Case study]. Praha: Karolinum.

- Fisher, R. (2011): *Učíme děti myslet a učit se* [Teaching children to learn]. Praha: Portál, 3. vyd.
- Hansen Čechová, B. (2009): *Nápady pro rozvoj a hodnocení klíčových kompetencí žáků* [Ideas for development and assessment of key competencies]. Praha: Portál.
- Hrabal, V. st. & ml. (2002): *Diagnostika* [Diagnostics]. Praha: Karolinum.
- Hübelová, D., Janík, T. & Najvar, P. (2008): Pohledy na výuku zeměpisu na 2. stupni základní školy: souhrnné výsledky CPV videostudie zeměpisu [Views on geography education on second grade at elementary school: summary CPV videostudy of geography]. *Orbis scholae*, 2(1), 53-72.
- Chrástka, M. (1999): *Didaktické testy* [Didactic tests]. Brno: Paido.
- Janík, T., Janíková, M., Najvar, P. & Najvarová, V. (2008): Pohledy na výuku fyziky na 2. stupni základní školy: souhrnné výsledky CPV videostudie fyziky [Views on physics education on second grade at elementary school: summary CPV videostudy of physics]. *Orbis scholae*, 2(1), 29-52.
- Kalhous, Z. & Obst, O. et al. (2002): *Školní didaktika* [School didactics]. Praha: Portál.
- Kolář, Z. & Šikulová, R. (2009): *Hodnocení žáků* [Student assessment]. Praha: Grada, 2. doplněné vyd.
- Kolář, Z. & Vališová, A. (2009): *Analýza vyučování* [Analysis of teaching]. Praha: Grada.
- Kosíková, V. (2011): *Psychologie ve vzdělávání a její psychodidaktické aspekty* [Psychology in education and its psycho-diagnostic aspects]. Praha: Grada.
- Košťálová, H., Miková, Š. & Stang, J. (2008): *Školní hodnocení žáků a studentů se zaměřením na slovní hodnocení* [School assessment of pupils and students with focus on word assessment]. Praha: Portál.
- Kratochvílová, J. (2011): *Systém hodnocení a sebehodnocení žáků* [System assessment and self-assessment of students]. Brno: MSD.
- Kyriacou, Ch. (2012): *Klíčové dovednosti učitele* [Essential Teaching Skills]. Praha: Portál, 5. vyd.
- Mandíková, D. & Palečková, J. (2007): Videostudie TIMSS 1999 – jak se vyučuje přírodním vědám v různých zemích [Videostudy TIMMS 1999 – how are natural sciences taught in different countries]. *Pedagogika*, vol. LVII, 238-250.
- Pasch, M. et al. (2005): *Od vzdělávacího programu k vyučovací hodině* [Teaching as decision making]. Praha: Portál, 2. vyd.
- Petty, G. (2008): *Moderní vyučování* [Teaching Today]. Praha: Portál, 5. vyd.
- Santiago, P., et al. (2012): *OECD Reviews of Assessment and Assessment in Education: Czech Republic 2012*, OECD Publishing. Available on: <<http://dx.doi.org/10.1787//9789264116788-en>>.
- Seebauerová, R., Helus, Z. & Kolidias, E. (ed.) (2000): *Kvalita cestou kvalifikace* [Quality through qualification]. Brno: Paido.
- Skalková, J. (2007): *Obecná didaktika* [General didactics]. Praha: Grada.
- Slavík, J. (1999): *Hodnocení v současné škole* [Assessment in the present school]. Praha: Portál.
- Starý, K. et al. (2008): *Učitelé učitelů: náměty na vzdělávání vlastního učitelského sboru* [Teachers' teachers]. Praha: Portál.
- Starý, K. et al. (2008): *Pedagogika ve škole* [Pedagogy in school]. Praha: Portál.

- Šeďová, K., Švařiček, R. & Šalamounová, Z. (2012): *Komunikace ve školní třídě* [Communication in school classroom]. Praha: Portál.
- Šimoník, O. (2005): *Úvod do didaktiky základní školy* [Introduction to elementary school didactics]. Brno: MSD.
- Vališová, A., Kasíková, H. (eds.) (2011): *Pedagogika pro učitele* [Pedagogy for teachers]. Praha: Grada, 2. rozšířené a aktualizované vydání.

2.3 Denmark: Review and syntheses of Danish research results on assessment – Danish LWG

Jan Alexis Nielsen⁵

2.3.1 Answers to the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

We lack research to document the “role [that] summative and formative assessment play in and for the teaching and learning” in Denmark. Thus, when answering this question, we can only describe how summative and formative assessment *shapes the context* of teaching and learning in Denmark at the different levels. You could add, that we have very little research in the area of assessment within STM. We do have substantial experience with experimental teaching involving new forms of assessment and evaluation. But these experiences have normally been recorded in reports in Danish and not published as peer reviewed research. There are some minor studies in mathematics education, however, that have pointed out that the heavy summative assessment practice in mathematics has an outspoken effect on the form and content of day-to-day teaching (Jess, 2005).

In Denmark there is a significant difference between the compulsory and upper secondary levels, therefore we have divided our answer to this question accordingly.

Summative assessment at the compulsory level

At the compulsory level, students – at the end of grade nine – have to take seven exams. Five of these exams are fixed (hereunder an exam that measures mathematical skills and mathematical problem-solving competences). Further, a centralized lot is drawn to select two additional exams – one in a subject from the humanities, and one in a subject from the natural sciences. All written exams are nationwide and are designed by a department of the Ministry of Children and Education.

All oral exams are school-based and are designed by the individual teacher. The Ministry of Children and Education publishes teacher guidelines for these exams.

Most written exams in the natural sciences are paper-pencil semi-open formats. Some items on the mathematics exam in compulsory school can be multiple-choice. In many other written exams such as Danish composition, students are allowed to use a computer.

The written and oral exams yield results at the individual level – although the Ministry of Children and Education perform some nationwide statistical analysis on the written exam results. The written and oral exams measure literacy, knowledge, skills and competences.

⁵ Corresponding author (e-mail: janielsen@ind.ku.dk)

Most written exams are graded by a selected ‘readers’ or ‘censors’ – i.e. teachers in the specific subject. All student responses are collected and graded by two independent teachers at a convention. Some of the written exams are graded by the class’ teacher and a second ‘reader’ or ‘censor’. Students rarely receive feedback on their performance. The oral exams are graded immediately after the oral examination by the examiner (the class’ teacher) and an independent second ‘reader’ or ‘censor’. Students receive – along with their grade – some feedback on their performance.

Crucially, none of these exams have direct consequences for the students. Teachers in the compulsory level will deem each students either ‘eligible’, ‘possible eligible’, or ‘not eligible’ to pursue an upper secondary education. The decision for each student will depend on the teachers’ long-term assessment of the student – not on the final exams.

In that sense, apart from mathematics, the centralized summative assessment at the compulsory level can be hypothesised to have a relatively low impact on the day-to-day teaching and the learning that takes place in it. However, it is our impression, that many teachers regularly use self-designed tests in their day-to-day teaching – tests which may or may not be systematically used for formative purposes.

Formative assessment at the compulsory level

In compulsory school, students must take 10 national tests that measure knowledge and skills in different subjects. These tests yield results at the individual, class, school, regional, and national level. The national tests in the compulsory level are nationwide and are designed by a department of the Ministry of Children and Education. The tests are computer-based and consist of multiple-choice and some open items. Further, the tests are computer analysed and are used primarily for benchmarking purposes and formative assessment of the student. The students can receive feedback on their individual performance.

While teachers are urged to use the test-information of the individual student in future planning of the teaching, there has been no research-based evaluation of how and to which extent teachers use these tests for formative purposes.

Summative assessment at the upper secondary level

There are four different strands of upper secondary schools in Denmark: (a) The general upper secondary schools; (b) the technical upper secondary schools; (c) the mercantile upper secondary schools; and (d) the higher preparatory upper secondary schools.

In school strand (a) through (c), the student must take 9 exams (oral plus written). There are two mandatory exams (one larger interdisciplinary project (“studieretningsprojekt”) and an exam in the subject “general study preparation” (“almen studieforberedelse”). Nationally drawn lots determine the subjects for the rest of the exams. (In some circumstances students have to take more exams. For example, if a student has opted to participate in an additional high-level subject, that student must be exam-

ined in that subject.). In school strand (d), the student ends all subjects with a final exam.

All written exams are nationwide and are designed by a department of the Ministry of Children and Education. And all oral exams are school-based and are designed by the individual teacher. The Ministry of Children and Education publishes teacher guidelines for these exams. Most written exams in the natural sciences are paper-pencil semi-open formats.

The written and oral exams yield results at the individual level – although the Ministry of Children and Education perform some nationwide statistical analysis on the written exam results. The written and oral exams measure literacy, knowledge, skills and competences.

Most written exams (both levels) are graded by a selected ‘readers’ or ‘censors’ – i.e. teachers in the specific subject. All student responses are collected and graded by two independent teachers at a convention. In the compulsory level, some written exams are graded by the class’ teacher and a second ‘reader’ or ‘censor’. Students rarely receive feedback on their performance. The oral exams are graded immediately after the oral examination by the examiner (the class’ teacher) and an independent second ‘reader’ or ‘censor’. Students receive – along with their grade – some feedback on their performance

In the upper secondary level, students will receive grades that reflect their performance in the day-to-day teaching in each subject. Further, the student will receive a grade for each written and oral exam. The weighted grade point average will determine which tertiary study programmes a student is eligible to.

The summative assessment at the upper secondary level has an outspoken impact on the day-to-day teaching and the learning that results from it, because the final grade point average determines the possible choices of education for each student, and because every subject can be drawn as a final-exam-subject, the summative assessment.

Formative assessment at the upper secondary level

The yearly grades given to students based on their day-to-day performance are summative, but they must be followed with an explanation/justification of the assessment. In other words, there are definite opportunities for ‘long-ranged’ formative assessment in upper secondary school.

We hypothesise that many teachers in upper secondary school are aware of a multitude of ‘shorter-ranged’ formative assessment methods – e.g. formative assessment within a shorter activity. Further, we hypothesise that while some teachers manage to positively shape the day-to-day teaching because of their use of such formative assessment, it is the case that the summative assessment has a much stronger influence on the day-to-day teaching. But these hypotheses remain unchecked.

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent does they use structured formative assessment and in what formats?

There is almost no research-based knowledge about this question in Denmark. However, a recent study that attempted to validate the PISA instrument, explored the evaluation culture in science subjects at the compulsory level (Dolin and Krogh, 2008). They found that when teachers assess their students, they have “an outspoken focus on learning and learning potential” and that most teachers assess “continuously *and* after the individual activity” (Dolin and Krogh, 2008, p. 69; our translation). The most common assessment methods were individual written tests and informal conversations with the whole class, while individual oral assessment was more uncommon.

A recent study showed that compulsory teachers were ambivalent towards the introduction of standardised tests (Schou, Hvidberg, and Kvols, 2009). On the one hand teachers welcomed the focus on assessment in compulsory school. On the other hand it left them with a fear of increased centralised management of their teaching. On the basis of the study, it was concluded that the introduction of standardised tests could be perceived as a “curtailment of the educational latitude of teachers” (Schou et al., 2009, p. 22; our translation).

Thus in compulsory science it does indeed seem (i) that teachers are aware of the need to continuously monitor student learning, (ii) that teachers generally attempt to cover this need in practice, and (iii) that teachers perceive threats towards their practice of continuously monitoring student learning.

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

There is a lack of research-based knowledge about this question in Denmark. Thus we can primarily sketch some of the systemic contexts that have impact on this issue.

From our perspective there is an outspoken need for pre- and in-service education about formative assessment for upper secondary school teachers. Upper secondary school teachers in Denmark are first and foremost academics with a major and minor subject. It is only after their education at university, that they encounter pedagogical and educational training. Recently most Danish universities have begun to more explicitly and systematically offer subject specific educational courses, but the course loads are still rather minimal (between 7.5 and 15 ECTS). Further, there is not much room for thematising formative assessment in the candidate teachers’ in-service teacher education program.

Since the publication of ministerial report in 2002, there has been a national focus on *competence-based* teaching (Niss and Jensen, 2002). Instead of having student knowledge or skills as goals, teaching at both compulsory and upper secondary level is

directed towards fostering students' competences. This has created a need for new ways and tools for assessment.⁶

A 2004 ministerial report on science teaching has argued that in order for good practice assessment in science to work, there must be (i) a stronger focus on assessment from both teachers and school leaders, (ii) a stronger relation between educational research and assessment practice, and (iii) an increased focus on formative assessment embedded in collaborative efforts at the level of teachers, institutions and researchers (Busch, Elf, and Horst, 2004). A 2006 ministerial report on mathematics teaching has argued that compulsory mathematics teachers need to be offered a wide range of formative assessment methods and tools, and that formative assessment cannot only be made in the form of standardised tests (such as the national tests) (Niss et al., 2006).

A study of the introduction of new forms of assessment in higher education mathematics courses pointed to the fact changing assessment formats requires that teachers and students address changes to the *didactical contract* – the typically implicit agreement on how mutual responsibilities are distributed between teacher and students (Grønbæk, Misfeldt, and Winsløw, 2009). In that sense the change of assessment practice at the level of classroom teaching in all levels, may be hypothesised to require renegotiating the didactical contract.

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results formatively and/or vice versa?

To our knowledge, there is no research-based knowledge about this question in Denmark. But there are explicit systemic aspects that ought to afford the use of summative assessment to support formative assessment.

In compulsory school, students must take 10 national tests that measure knowledge and skills in different subjects. Teachers are urged to use the test-information of the individual student in future planning of the teaching.

In upper secondary school, the grades given to students based on their day-to-day performance are summative, but they must be followed with an explanation/justification of the assessment, normally as a person to person consultation between the teacher and the individual student. These grades are given three-four times a year in each subject and all but the last given grade can thus be seen as a formative feedback. The final grade in a subject can also be seen as a formative assessment for those students who are continuing the subject at a higher level (most subjects are offered at three levels, level C, B, A, the higher levels have as a necessary prerequisite that you have taken the lower levels). In other words, there are definite opportunities for 'long-ranged' formative assessment in upper secondary school.

⁶ See e.g. Højgaard, T., Sølberg, J., Bundsgaard, J., Elmose, S., Kompetencemål i praksis – foranalysen bag projektet KOMPIS. (English title: Competence goals in practice – a pre-analysis behind the project KOMPIS). In *MONA*, 2010(3), 7-29, 2009.

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

See question 3 above.

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

To our knowledge, there is no research-based knowledge about whether “assessment methods influence the uptake of IBE” in Denmark. But from our experience from talking to teachers in upper secondary school, we hypothesise that most teaching in science subjects is geared towards the final examinations (that are not designed to measure inquiry-competences). Accordingly, we hypothesise that often *when* IBE takes places it is perceived as auxiliary to the ‘core’ teaching geared towards the examinations.

There is some research on how “summative and formative assessment methods [can] be used to promote learning” in IBE.

In his doctoral dissertation, Jens Dolin (2002) shows how an authentic physics teaching (which to a large degree is identical with an IBE approach) enhances student learning and motivation. Such a teaching also has the effect that a larger proportion of the students, and especially the female students, choose to take the subject at a higher level.

In his dissertation, Christensen (2004) designed an “integrated assessment” method that ought to be fruitful for promoting learning in IBE. The method is called “The Teacher-Student Assessment Dialogue on Subject Matter” and attempts to “integrate assessment and instruction in one and the same Teacher-Student Dialogue” (p. 275). The tool consists of a physical arrangement of the students and a fixed structure of the dialogue between one student and the teacher, which was observed by the rest of the class and afterwards discussed in plenum. On the basis of his study, he formulated 10 principles for an “ideal type of the Assessment Dialogue” (Christensen, 2004, p. 282; see attachment). In his study Christensen documented that the integrated assessment method is conducive to student learning and can amplify the learning processes that are prompted by regular teaching; both effects hold true for the student being assessed and the students in the audience (Christensen, 2004, p. 246). Further, teachers and students who participated generally reported (among other things) that the method provides shared learning instances that teachers and students can draw on in future teaching, and that the method is a means for students’ self-evaluation (Christensen, 2004, p. 261)

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

Dolin (2002) includes an action research project with three physics teachers together with a researcher developing an assessment design able to capture process competences in physics, especially competences attached to practical work. The assessment method was developed continuously through the two years of compulsory physics at the science stream in upper secondary and used formatively to teach students to perform inquiry processes. It consisted of a template, ending with eight categories of both subject specific character (such as the ability to control variables) and more generic character (such as ability to work together). The template was used in the final examination where the students in groups of two-three were performing an investigation of an unknown phenomenon through a four hour process. The teacher and two censors were following the students and grading them individually. The reliability was of the same order as at normal oral examinations (0.90). The students' motivation were dramatically increased – especially the girls'. The project also established close connections between the science teachers' approach to teaching (measured via the ATI instrument) and his or her readiness to engage in an authentic science teaching.

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

Besides the methods devised by Dolin (2002) and Christensen (2004) that were mentioned above, a host of instruments and protocols exist. *But there is very little research-based knowledge about "which and how are they used"*.

The Ministry of Children and Education have published guidelines and instructions for a total of 25 different tools or methods for assessment – most of them can be used directly for formative assessment. Many of the instruments are described in detail with e.g. (i) rationale for using the instrument, (ii) concrete guides to use the instrument, (iii) affordances and drawbacks of the instrument; and most descriptions are supplemented with concrete examples. In other words, it should be relatively easy for teachers to make informed decisions about whether and when to use a specific instrument – albeit actually implementing the instruments still must require experience. We have listed the most notable instruments here:

- *Action learning* (formative plus summative instrument modelled on action based research, where the teacher (a) identifies a problem, (b) designs and implements an action, (c) observes, (d) reflects, and (e) creates a new action). [Generic description with an example]
- *The Delphi-method* (formative and/or summative instrument much like how the method is used in research) [Fairly detailed description with a few examples]

- *Student logs* (formative instrument which the teacher can continuously tap into for information about student progress) [Fairly rich description with examples]
- *Assessment dialogues* (formative and summative instrument very similar to the method devised by Christensen (2004)) [Rich description with examples]
- *Students' storytelling* (formative instrument geared to flesh out and make concrete one or more aspects that influences an individual students' learning and/or wellbeing). [Fairly rich description]
- *Day-to-day assessment* (generic type of formative assessment, typically with focus on self-assessment, geared towards including students in e.g. the planning of teaching) [Generic description]
- *Can – Almost can* (formative assessment instrument designed to help teachers and students realize what a student can, almost can, or cannot, typically used for making longer term student plans) [Rich description with examples]
- *Classroom parliament* (formative and summative instrument designed to let teachers and students collaboratively identify barriers for student learning) [Rich description with examples]
- *The quality star* (formative instrument designed to organise a mutual agreement between teacher and students about what the aims of a given teaching activity is. Further the instrument is designed to foster alignment between the learning goals and the teaching). [Fairly rich description with one example]
- *The quality staircase* (similar to 'The quality star' but with a heavier focus on learning progression within one or more activities) [Fairly rich description with one example]
- *Checking by the teacher* (generic formative and/or summative assessment in which the teacher dialogically checks either understanding or learning progression of individual students) [Generic description]
- *Teacher logs* (formative assessment instrument that allows the teacher to structure her reflections about teaching for the purpose of designing future teaching) [Fairly rich description with one example]
- *Natural development* (formative and/or summative assessment instrument designed to capture aspects that are difficult to systematise or aspects that are not planned in teaching. In which a student tells about how she experienced what happened in a given situation. This leads to a reflection and a dialogue about what the student learned from the experience) [Fairly rich description]
- *Portfolio* (similar to the approach used in many other countries) [Rich description with a number of examples]
- *Collaboration contract* (generic assessment instrument designed to allow multiple actors to adjust expectations and make binding agreements. Can be used formatively) [Rich description with examples]
- *SMTTE – Context, Goal, Signs, Actions, Assessment* (generic assessment and planning instrument designed to plan, implement and assess teaching) [Rich description with examples]
- *Time registration* (formative and/or summative assessment instrument in which students keep track of the time spent of various types of activities. In parallel they annotate (qualitatively) the time registration sheet) [Fairly rich description]

- *The developmental spiral* (Formative assessment instrument as well as planning instrument designed make concrete learning goals secure alignment between goals and activities) [Generic description with example]
- *Observation of teaching* (Generic formative and/or summative assessment instrument) [Rich description of various techniques]

Further some literature exists that provide tools for assessment for mathematics teachers, in particular.⁷

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

Unfortunately there are only rare instances in which research on assessment is used by other than researchers.

Question 10: Which are the 10 most relevant publications on formative and summative assessment of STM in your country?

Christensen, T. S. (2004): *Integreret Evaluering – En undersøgelse af den fagligt evaluerende lærer-elevsamtale som evalueringsredskab i Gymnasial Undervisning.* (English title: *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. Retrieved from http://static.sdu.dk/mediafiles/Files/Om_SDU/Fakulteterne/Humaniora/Phd/afhandlinger/2005/Afhandlinger%2042_spanget%20pdf.pdf

Davidsson, E. Sørensen, H., & Allerup, P. (2012): *Assessing scientific literacy through computer-based tests – consequences related to content and gender.* In *Nordic Studies in Science Education*, 8(3), 269-282.

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

Dolin, J., & Krogh, L. (2008): *Den naturfaglige evalueringskultur i folkeskolen* (English title: *The culture of assessment in compulsory school science*). In *MONA*, 2008(4), 60-69.

Dolin, J., & Krogh, L. (2010): *The Relevance and Consequences of PISA Science in a Danish Context.* In *International Journal of Science and Mathematics Education*, 8(3), 565-92.

Grønbæk, N., Misfeldt, M. & Winsløw, C. (2009): *Assessment and Contract-Like Relationships in Undergraduate Mathematics Education.* In O. Skovsmose, P. Vale-

⁷ e.g. Dahl Søndergaard, B., Lilholt, A., Olesen, A., Skipper-Jørgensen, A. & Wahl Andersen, M., *Evaluering og test i matematik.* (English title: *Assessment and test in mathematics*). Kroghs Forlag, 2007.

- ro, & O. Ravn Christensen (Eds.), *University Science and Mathematics Education in Transition*. New York: Springer, 85-105.
- Jess, K. (2004): *Formativ evaluering i matematikundervisningen - Ændringer i Praksis*. (English title: *Formative assessment in mathematics teaching*). *Nordic Studies in Mathematics Education*, 9(4), 15-47.
- Krogh, E., & Jensen, M. J. (2003): *Portfolioevaluering - en rapport om en ny evaluerings- og arbejdsform*. (English title: *Portfolio assessment – a report on a new way of assessing and working*). *Udviklingsprogrammet for fremtidens ungdomsuddannelser* – nr. 46. Retrieved from <http://www.emu.dk/gym/fag/st/evaluering/Portfolioevaluering.pdf>
- Miller, T. (2004): *Karaktergivning i praksis 13-skalaen og gymnasiet*. (English title: *Grade-giving in practice – the 13-grade scale and the upper secondary school*). Phd Dissertation, University of Southern Denmark. Retrieved from http://static.sdu.dk/mediafiles//Files/Om_SDU/Fakulteterne/Humaniora/Phd/afhandling/2004/Afhandling%2026_miller%20pdf.pdf
- Niss, M. (Ed.) (1992): *Cases of Assessment in Mathematics Education: An ICMI Study*. Dordrecht: Springer.
- Tanggaard, L., & Elmholdt, C. (2008): *Assessment In Practice: An inspiration from apprenticeship*. *Scandinavian Journal of Educational Research*, 52(1), 97-116.

Additional information from the authors

With respect to technology education, there is some form of integration of technics in science in the first grades but practically no research into assessment.

References

- Busch, H., Elf, N. F., Horst, S. (2004): *Fremtidens Uddannelser – Den nye faglighed og dens forudsætninger*. (English title: *The education of the future – The new disciplinarity and its preconditions*). Ministry of Education.
- Christensen, T. S. (2004): *Integreret Evaluering – En undersøgelse af den fagligt evaluerende lærer-elevsamtale som evalueringsredskab i Gymnasial Undervisning*. (English title: *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. Retrieved from http://static.sdu.dk/mediafiles/Files/Om_SDU/Fakulteterne/Humaniora/Phd/afhandling/2005/Afhandling%2042_spanget%20pdf.pdf
- Dolin, J. (2002): *Fysikfaget i forandring – læring og undervisning i fysik i gymnasiet med fokus på dialogiske processer, autenticitet og kompetenceudvikling* (English title: *Physics in school under change. Learning and teaching in physics in upper secondary with focus on dialogical processes, authenticity and competence development. With an English summary*). Phd Dissertation, Roskilde University. Retrieved from <http://rudar.ruc.dk/handle/1800/1645>.
- Dolin, J., & Krogh, L. (2008): *Den naturfaglige evalueringskultur i folkeskolen* (English title: *The culture of assessment in compulsory school science. With an English*

- summary). Series of the Department of Science Education 2008(17). Retrieved from http://www.ind.ku.dk/publikationer/inds_skriftserie/vap-rapport_-_den_naturfaglige_evalueringskultur_i_folkeskolen_/indskriftserie17vap-webversion.pdf/
- Grønbæk, N., Misfeldt, M. & Winsløw, C. (2009): Assessment and Contract-Like Relationships in Undergraduate Mathematics Education. In O. Skovsmose, P. Valero, & O. Ravn Christensen (Eds.), *University Science and Mathematics Education in Transition*. New York: Springer, 85-105.
- Jess, K. (2009): Konsekvenser af evaluering i matematikundervisning. (English title: Consequences of assessment in mathematics teaching). In *MONA*, 2005(2), 22-39, 2005 & Svendsen, J., Matematiklærerens forberedelse. (English title: The preparation of the mathematics teacher) *INDs studenterserie nr. 9*, 2009.
- Ministry of Children and Education (no date): *Værktøjer til evaluering i folkeskolen*. Available on <http://uvm.dk/Uddannelser-og-dagtilbud/Folkeskolen/De-nationale-test-og-evaluering/Evaluering/Vaerktoejer>.
- Niss, M., & Jensen, T.H. (2002): Kompetencer og matematiklæring. (English title: Competences and the learning of mathematics). Ministry of Education.
- Niss, M. et al.(2006): Fremtidens matematik i folkeskolen. (English title: Future mathematics in the compulsory school). Ministry of Education.
- Schou, L. R., Hvidberg, C., & Kvols, A. M. (2009): Standardiserede test og evalueringer i folkeskolen. (English title: Standardized tests and assessment in compulsory school). In *Bedre skole*, 2009(2), 19-23.

2.4 Finland: Report from Finland concerning formative and summative assessment with respect to IBE and STEM

Pasi Nieminen⁸ and Jouni Viiri (University of Jyväskylä)

2.4.1 Foreword

We tried to find appropriate research from databases of Finnish universities and national and international journals. We also contacted many persons who have worked as researchers and/or have acted in the Finnish National Board of Education or the Finnish Institute for Educational Research. Unfortunately, it seems that there are only few small-scale studies conducted in Finland concerning the topic. There are some large-scale studies related to summative assessment, such as studies of the Finnish National Board of Education or PISA and TIMSS studies. However, these are not really studies *about* summative assessment, but they are summative assessment about students' skills in mathematics and science. Nevertheless, they may give some hints concerning classroom assessment in Finland.

Thus, we are not able to give *research-based* answers to all questions presented in the guidelines document. We can answer, more or less, to the questions 1 and 10: a general picture based on few studies about formative and summative assessment in Finnish STEM classes and "most relevant" publications. We try answer hypothetically (not research-based) the questions 2-9.

2.4.2 Answers to the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

Summative assessment in STEM

Undoubtedly, summative assessment has central role in Finnish classrooms. Although we have no high-stakes testing, students face lots of exams composed by teachers. Lower secondary students study about twenty different subjects and most of them include one or more summative tests during a semester. Hence, students have been tested about 50 times per year. (Atjonen, 2007; Kupiainen, Hautamäki, & Karjalainen, 2009) Results of these teacher generated summative exams are probably the biggest factor when teachers give grades to students in STEM.

Formative assessment in STEM

It is difficult to say how much formative assessment appears in Finnish STEM classes, because we have no large-scale studies for that question. There are some statements concerning formative assessment in the National curriculum (Finnish National Board of Education, 2004), but these do not guarantee for what is happening in reality in classrooms. For example, the curriculum gives criteria for grade 8 (summative grades are given from 4 to 10), but still there are significant differences between schools what is demanded for grade 8 (Kuusela, 2006).

⁸ Corresponding author (e-mail: pasi.k.nieminen@jyu.fi)

Some large-scale studies give hints about (lack) of formative assessment. For example, a study (169 schools, 6555 students of 9th grade) of the Finnish National Board of Education showed that a frequency of tests and exams is between “sometimes” or “often” (teachers’ and students’ opinions on likert scale questions). In contrast, “Students set their own targets and evaluate their progress” happens less frequently than “sometimes” (Rautopuro, 2013). Also in 9th grade physics, chemistry, biology and geography lesson “Students set their own targets and evaluate their progress” less frequently than “sometimes” (Kärnä, Hakonen, and Kuusela, 2012).

An OECD report (Organisation for Economic Co-operation and Development, 2005) attempts to describe formative assessment practices in two Finnish lower secondary schools in Jyväskylä (Tikkakoski) and Helsinki (Meilahti). As the report is based on observations, it may have some valid findings. However, it is only a small-scale study which results are not very generalizable.

The report describes mathematics lessons in Tikkakoski: first a teacher gives short lecture (10-15 min) about topic of lesson, and then students are working with mathematical problems alone or with a pair. Students can ask questions (also during a short lecture) and teacher walks around and gives help. This description sounds very familiar to us (authors) as former lower secondary teachers. We think that typical lessons may include some moments which are favorable for formative assessment (students’ questions, peer discussions and teacher’s personal help).

On the other hand, Lehesvuori et al. (accepted) analysed classroom discussion in lower secondary physics classes (25 double lessons, ninth grade). They found that only three of the 25 teachers discussed with students in an interactive dialogic way wherein students’ own conceptions were probed. Hence, we think that typically the classroom discussion in Finnish physics classrooms is not very supportive for formative assessment as, for example, Ruiz-Primo (2011) has described such kind of discussion.

It is stated in the mentioned OECD report (Organisation for Economic Co-operation and Development, 2005) that the math teachers “use frequent short tests - once a week - to see what problems students have understood”. We think that these short tests may have formative function. However, it must be noted that once a week sounds very frequent, and it cannot be deduced that all Finnish teachers use short tests so often. Further, teachers may also take short tests into account in summative grading.

Further, the OECD report mentions the student self-assessment at the end of period (in Tikkakoski five times per year). When the author Pasi Nieminen was a lower secondary teacher in 2008-2009, students in his school made self-assessment (for all subjects) two times per year at the end of a semester. We are not sure how effective this kind of self-assessment is as formative assessment method, because it is done *after* a course or a semester.

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?

We think that formative assessment mostly occurs in classroom interaction (e.g., a discussion between a teacher and a student). Probably, this kind of interaction depends on a teacher and students, and then its occurrence varies between classes. It is also a standard Finnish method to check the homework (revision) at the beginning of the lesson. Teachers check homework assignments: e.g., they check how students have succeeded to solve problems and decide if some assignments are needed to discuss more closely. In addition, they may questions about the most important topics of the previous lesson. We think this could be taken as formative assessment.

Presumably, structured formative assessment methods are quite uncommon in Finland. As mentioned before, teachers use short tests in mathematics classes which may have formative function.

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

We think that in Finland teachers understand the importance of formative assessment, but it is difficult to implement in daily practice. Hence, teachers need explicit examples and training for FA methods.

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results formatively and/or vice versa?

We think that summative exams may have formative effect in the long term, because a teacher may change his/her lesson plans if it seems that students have not learnt some issues. Students may profit formatively from summative exams as well. However, we suppose that a student's (meta) learning skills have an effect to how much he/she can benefit of summative assessment (when he/she see results of an summative exam). For example, many lower secondary students do not care to reflect their results after summative exam. They just look how many points they have received in the exam. On the other hand, summative assessment (grades) may have detrimental effect on student learning. A student may start to think that he/she is "a low level student" which may have an influence to his/her effort.

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

We think that Finnish teachers understand the importance of formative assessment. The biggest challenge might be that formative assessment can be difficult and laborious to implement. Formative assessment should be promoted in pre-service training, as in Finland we have no obligatory in-service training having some continuous pedagogical development (only three days per year and this might include almost whatever related to teaching). One way is to discuss with textbook writers, as textbooks have very central role in Finnish teaching and learning.

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

This is very difficult question to answer as we do not know how much IBE we have in Finnish science and mathematics classes. We think that pure IBE is quite uncommon. In science classes we have lot of practical work which may have some features of IBE. However, a teacher mostly poses research questions of inquiry. Hence, "inquiry" is quite teacher guided.

We think that our assessment do not support IBE at the moment. Students' grades are mostly determined using summative exams which measure other skills than inquiry skills.

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

We have not found research about this. As stated in previous question, we think IBE is quite uncommon. If a teacher would use IBE, he/she should assess students (also summatively) when they are doing inquiry. Paper-and-pencil exams (which are used in Finnish classes) are not appropriate to assess inquiry skills, but they should be evaluated in real situations. Of course, inquiry skills should be taken into consideration in summative grades if IBE would be used.

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

No.

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

It is difficult to give even a hypothetical answer as there are so few studies concerning assessment.

Question 10: Which are the 10 most relevant publications on formative and summative as-sessment of inquiry in science, technology, and mathematics in your country?

Formative assessment

- Organisation for Economic Co-operation and Development (2005): *Formative assessment: Improving learning in secondary classrooms* OECD publishing. First, the report describes assessment in Finnish schools rest on the principles of the National Board of Education. In addition, the report presents results of observations in two Finnish lower secondary schools (see text above).
- The data of study Lehesvuori et al. (accepted). The paper does not concern formative assessment, but the video data is good and large and we have an access on it. We have think that we could analyse it from perspective of formative assessment.

Formative assessment & inquiry

- Ahtineva, A. (2004): Laboratory assessment in chemistry education. In A. Laine, J. Lavonen, & V. Meisalo (Eds.), *Current research on mathematics and science education: Proceedings of the XXI annual symposium of the Finnish Association of Mathematics and Science Education Research* (252-264).
<http://www.edu.helsinki.fi/malu/tutkimus/tutkimusseura/proceedings2004.pdf>
The paper presents small-scale study in lower and upper secondary chemistry. "Lab licence" card for 7.th grade students, observation card for 8.th grade students' laboratory working skills and evaluation of upper secondary students' ($n = 16$) inquiry.

Summative assessment

- Tikkanen, Greta, and Aksela, Maija (2012): Analysis of Finnish chemistry Matriculation Examination questions according to Cognitive Complexity. *Nordic Studies in Science Education* 8(3), 257-268.
<https://www.journals.uio.no/index.php/nordina/article/view/532/578>
The paper presents an analysis of Finnish chemistry matriculation examination questions according to cognitive complexity (257 questions from 28 matriculation examinations between 1996 and 2009). Qualitative approach and theory-driven content analysis method using Bloom's revised Taxonomy of Cognitive Objectives were employed in the research. The research indicates that the examinations were cognitively demanding. According Tikkanen et al., the Bloom's revised Taxonomy of Cognitive Objectives as used in this research gives a useful way for designing or analysing chemistry summative assessment tools. Matriculation examination has very central role in Finnish upper secondary education as its results are used in selections for tertiary education. Hence, it has a big influence on what is taught in upper secondary schools.

Inquiry

- Hähkiöniemi, M., Leppäaho, H., & Francisco, J. (in press): Teacher-assisted open problem-solving. *Nordic Studies in Mathematics Education*.
Abstract: Previous research has developed several problem-solving models and suggested that the teacher plays a crucial role in guiding students' problem solving. However, less is known about the particularities of problem solving and teacher guidance when dealing with open problems which include multiple possible solution pathways. The aim of this study is to understand students' open problem-solving processes and teachers' ways of supporting them. Data collection involved videotaping one 9th grade mathematics lesson with two video cameras and capturing the screens of the students' computers. Seven student pairs worked on an open problem using GeoGebra under the guidance of a teacher trainee. We found that students had various kinds of problem-solving processes and that the teacher had a crucial role in guiding them. We elaborate on 9 ways how the teacher guided students to change between phases in open problem-solving.
- Kim, M., Lavonen, J., Juuti, K., Holbrook, J., & Rannikmäe, M. (2012): Teacher's reflection of inquiry teaching in Finland before and during an in-service pro-

gram: Examination by a progress model of collaborative reflection. *International Journal of Science and Mathematics Education*, 11(2), 359-383.

Abstract. In inquiry-based science education, there have been gradual shifts in research interests: the nature of scientific method, the debates on the effects of inquiry learning, and, recently, inquiry teaching. However, many in-service programs for inquiry teaching have reported inconsistent results due to the static view of classroom inquiries and due to the partial perspective between individual and collaborative reflections. Thus, by means of a theoretical progress model of collaborative reflection, this qualitative research aims to investigate reflections of four participant teachers before and during a half-year in-service teacher program. The model captures the following four interactions for each individual teacher and among the teacher cohort: belief to practice, practice to belief, stimulation, and reinforcement. The audio-video data and their quantification allowed identification of the teachers' consistent prior beliefs and practices as a multiplicity of inquiry teaching and their interwoven progress during the program. The findings are further discussed in terms of the *implicit development* and the *richer repertoire*.

Additional information from the authors

In Finland there is no such technology education (science-technology-society) as in some countries. There is "craft" and "knitting" education (some in Finland call that technology education). It is not connected to science and mathematics education.

References

- Atjonen, P. (2007): *Hyvä, paha arviointi*. Helsinki: Tammi.
- Finnish National Board of Education. (2004): National core curriculum for basic education 2004. Retrieved from http://www.oph.fi/download/47674_core_curricula_basic_education_5.pdf
- Hähkiöniemi, M., Leppäaho, H., & Francisco, J. (in press): Teacher-assisted open problem-solving. *Nordic Studies in Mathematics Education*,
- Kärnä, P., Hakonen, R., & Kuusela, J. (2012): *Luonnontieteellinen osaaminen perusopetuksen 9. luokalla 2011* Opetushallitus.
- Kupiainen, S., Hautamäki, J., & Karjalainen, T. (2009): *The finnish education system and PISA*. Finland: Ministry of Education.
- Kuusela, J. (2006): *Temaattisia näkökulmia perusopetuksen tasa-arvoon* Opetushallitus.
- Lehesvuori, S., Viiri, J., Rasku-Puttonen, H., Moate, J., & Helaakoski, J. (accepted): Visualizing communication structures in science classrooms: tracing cumulativity in teacher-led whole class discussions. Accepted in *Journal of Research in Science Teaching*.
- Organisation for Economic Co-operation and Development. (2005): *Formative assessment: Improving learning in secondary classrooms* OECD publishing.
- Rautopuro, J. (Ed.). (2013): *Hyödyllinen pakkolasku: Matematiikan oppimistulokset peruskoulun päättövaiheessa 2012* Opetushallitus.

Ruiz-Primo, M. (2011): Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15-24.

2.5 France: Contribution of ICAR – CNRS, France

Florence Le Hebel⁹, Sylvie Coppé, Suzane El Hage, Pascale Montpied, Andrée Tiberghien, & Zeynab Badreddine

2.5.1 Answers so the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

We found in the literature relative to France three research reports giving some elements of answer to this question: Issaieva et al (2011), Talbot (2009), and Mercier-Brunel & Jorro (2009).

- Issaieva, E. I., Pini, G., & Crahay, M. (2011): Positionnements des enseignants et des élèves du primaire face à l'évaluation: une convergence existe-t-elle ? *Revue Française de Pédagogie*, 176, 5-26.

In this research, made in primary school, these authors propose an approach aimed at finding how teachers and students may share the same beliefs or conceptions about the practice of formative assessment.

For the authors, formative assessment is performed either to enable the teachers to adjust and help students to progress or “to generate a self-regulation of current learning by the students” (p.8). In contrast, summative assessment and normative assessment aim to identify the students according to their abilities or skills, or to punish and classify them.

The data are obtained from the observation of 50 French classrooms of CM2 during French and mathematics lessons (5th year of primary school) in 39 schools located in three districts. The sample includes 1112 students: 552 girls (49.6% of the total) and 560 boys (50.4% of the total).

The main results underline the way teachers and students dealing with assessment:

The teachers' way of dealing with assessment falls into three profiles:

- Profile A includes teachers favoring formative assessment (49%),
- Profile B includes teachers considering that the formative and the summative approaches are equally important (32%),
- Profile C is quite mixed, as it gathers teachers opposed to several aspects of certified qualifications (normative assessment), but do not seem to adhere more strongly to formative assessment (19%).

The students' positions regarding assessment fall into three profiles:

- In profile A, the students are in favor of formative assessment and opposed to certification assessment (50%),

⁹ Corresponding author (e-mail: Florence.Le-Hebel@ens-lyon.fr)

- In profile B, the students gave an equal importance to formative and normative approaches (40%),
- In profile C, the students are characterized by an average consensus about the formative assessment and a low rejection of the normative aspects of the assessment (10%).

The students are also interrogated on their teachers' views about assessment. The results obtained allow building the following table about the convergence and divergence between students and teachers views. It allows as well estimating how students' awareness and abilities to identify their teachers views of assessment.

Table 2.5.1: Classrooms in which the students' positions are either convergent or divergent with teachers' conceptions and in which the students correctly or not perceive their teachers point of view. (Issaieva et al., 2011, translation of table 9, p.15)

Teacher's Cluster	Classrooms for which the students' conceptions and their teacher's conceptions are convergent	Classrooms for which the students' conceptions and their teacher's conceptions are divergent	Classrooms for which the perception of the teacher's conception by the students is exact	Classrooms for which the perception of the teacher's conception by the students is inexact
A	6 classrooms (1, 3, 18, 23, 39, 47)	1 classroom (14)	7 classrooms (1, 2, 18, 23, 36, 47, 48)	4 classrooms (22, 35, 42, 45)
B	1 classroom (13)	2 classes (33, 37)	2 classrooms (11, 17)	5 classrooms (6, 13, 33, 37, 44)
C	No classroom	2 classrooms (15, 49)	No classroom	5 classrooms (12, 19, 38, 41, 49)
Total	7 classrooms	5 classrooms	9 classrooms	14 classrooms

In conclusion, in a few classes, at the primary level, there is convergence between teachers and students views about assessment. In addition, there are also few classes (8/47) where students correctly apprehend the posture of their teachers on assessment. More impressive, in 14 classes, students' perceptions of their teacher's views on assessment are different.

- Talbot, L. (2011): Les pratiques d'évaluation orale des enseignants du primaire et du secondaire. *Mesure et Évaluation en Éducation*, 34(3), 79-112. (More details in question 2)
- Mercier-Brunel, Y., & Jorro, A. (2009): La parole évaluative de l'enseignant. *Les Dossiers des Sciences de l'Éducation*, 22, 9-24. (More details in question 8)

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?

- Talbot, L. (2011): Les pratiques d'évaluation orale des enseignants du primaire et du secondaire. *Mesure et Évaluation en Éducation*, 34(3), 79-112

This paper deals with the study of the practices of teachers in relation with student learning strategies within the classroom. The authors focus in particular on how individual oral assessment is carried out. The aim is to describe, understand and explain the work of nine teachers, five in primary and four in secondary education. It is shown that their verbal assessment habits can be characterized by both intra and between individuals as well as by a certain number of stabilities. Through the use of video-taped sequences and statistical analyses, it is possible to make sense of these variations and stabilities by studying the correlation between the level of teaching, the level of education and the gender of the student population.

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

No appropriate research paper found.

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results used formatively and/or vice versa?

No appropriate research paper found.

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

No appropriate research paper found.

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

No appropriate research paper found.

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

No appropriate research paper found.

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

In this question we consider different types of instruments and protocols.

The first reference (Grugeon et al., 2012) relates the use of a software (1). In contrast, there are other cases where we consider instruments like:

- (2) Strategies structuring feedback (Georges & Pansu, 2011),
- (3) Descriptive typology proposed to the teachers (Jorro & Mercier-Brunel, 2011),
- (4) Exercises given to the student for preparing summative assessment (Antibi, 2007),
- (5) Material written support of exchange the “personal contract of success” (Talbot, 2009),
- (6) Database reporting on how a series of activities should be used (Ministère de l'Éducation Nationale, 2012).

- (1) 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*, 137-162.

In the case of ICT, we have only one case applied to mathematics. The authors' (Grugeon et al, 2012) developed a software (Pépité) that can be used typically as an instrument for diagnostic assessment.

The authors define diagnosis assessment as follow: “diagnosis assessment informs teachers on knowledge as well as skills or misconceptions of students. It allows the teacher to identify what students are able or not able to do compared to institutional expectations”

The diagnosis helps teachers to adjust the teaching progression to meet the need of the students.

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

The authors propose theoretical main categorizations of feedback considering as a component of the formative assessment and learning. The feedback was categorized according to several criteria. The feedbacks can have a positive or a negative nature, can be assertive and evaluative, and can be linked to simple or complex situations.

For assertive feedback only simple situation exists “it is right” / “it is wrong”. In contrast, for the evaluative feedback simple and complex situations exist, respectively “it's good”, “you made efforts” / “it's bad”, “you didn't make enough efforts”.

The aim of this descriptive instrument (the feedback) is to help students to recognize what they already master or not, and what is their progression.

- (3) Jorro, A., Mercier-Brunel, Y. (2011): Les gestes évaluatifs de l'enseignant dans une tâche de correction collective. *Mesure et Évaluation en Éducation*, 34(3), 27-50.

This article aims to characterize the evaluative acts of teachers and their impacts on students' learning during collective correction phases. Several postures based on enunciative indicators are defined:

- The normative posture,
- The « indicielle » posture oriented on a specific point,
- The « présenteielle » posture understandable only in regard to context where events reflects very active subjectivity but with little wording.

Based on the observations of classrooms, correspondences are built between teachers and students postures.

- (4) Antib, A. (2007): *Les notes : la fin du cauchemar ou en finir avec la constante Macabre*. Math'adore – Nathan, 158 p.

Antibi proposes an instrument in order to avoid that the distribution of the students' marks in the class corresponds to a Gaussian. He proposes to prepare the students to formal assessment by training students to typical exercises corresponding to that will be given to summative assessment.

- (5) Talbot, L. (2009): *L'évaluation formative : comment évaluer pour remédier aux difficultés d'apprentissage*. Paris: Armand Colin, 191 p.

From previous research studies, Talbot defines the assessment on the basis of several characteristics: the value, the measurement, the meaning, the evolution and the relation with the norms, and the judgment.

The author proposes an instrument called the "personal contract of success" which is independent of a specific discipline. This instrument is proposed to the student among those proposed by the teacher; it permits him/her to recognize his own difficulties; then the teacher prescribes activities of remediation to each student for his/her recognized difficulties; and at last the teacher checks if the students have correctly carried out these activities.

- (6) Ministère de l'éducation nationale. (last update October 2012). Banque de situations d'apprentissage et d'évaluation pour la compétence 3. Retrieved June 11, 2013, from Eduscol: <http://eduscol.education.fr/cid55510/banque-de-situations-d-apprentissage-competence-3.html>

Database of learning and assessment situations for one of the pillar of the common base of knowledge and skills: Basic knowledge in mathematics and scientific and technological culture.

This database is proposed by the French Ministry of education. It consists of a series of activities: 14 activities in mathematics, 6 activities in physics, 5 activities in biology and

geology, 2 activities on technology and 8 activities on interdisciplinary between mathematics and physics.

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

No appropriate research paper found.

Question 10: Which are the 10 most relevant publications on formative and summative assessment of inquiry in science, technology, and mathematics in your country (please make sure that your selection covers all three subject areas)? If there are no publications especially dealing with IBE, please list the 10 most relevant publications in formative and summative assessment in general. What are the main results of these publications?

We found 10 publications that have relevance to formative and summative assessment of inquiry in science, technology, and mathematics in France; these publications are however not typically dealing with IBE.

The following list presents the 10 references ordered from the most relevant to the less relevant publication. For each reference we give when possible the abstract written in English by the authors, or our translation of the authors' abstract written in French; in the case of the books we translated a short introduction, (or a summary) written by the authors, or a foreword (or a preamble) given by the editors.

- Issaieva, E. I., Pini, G., & Crahay, M. (2011): Positionnements des enseignants et des élèves du primaire face à l'évaluation: une convergence existe-t-elle? *Revue Française de Pédagogie*, 176, 5-26.

Cet article se propose d'étudier les positionnements des enseignants et des élèves face à l'évaluation, ainsi que la manière dont les élèves interprètent la posture de l'enseignant. Plus précisément, il s'agit d'examiner les variabilités interclasse et intra-classe en ce qui concerne les postures des enseignants et des élèves. L'objectif est d'identifier les classes dans lesquelles les acteurs scolaires ont des conceptions convergentes, l'hypothèse étant que, dans ce cas, un espace commun de conceptions évaluatives réunirait les élèves et leurs enseignants.

Dans la foulée, il s'agit également d'étudier dans quelles classes les élèves perçoivent les conceptions des enseignants de manière exacte. Les résultats mettent en évidence différents profils de positionnement chez les enseignants et les élèves ; ces profils se ressemblent fortement. Par ailleurs, on retrouve peu de classes à l'intérieur desquelles les points de vue des acteurs sont en concordance.

Here is the translation of the French abstract:

The purpose of this article is to study the teachers and students' positions on assessment, and how students interpret the teacher position. More specifically, the aim is to examine the variability of teachers and students postures both within the classroom

and between different classrooms. The aim of this article is to identify in which classrooms both actors, “students and teachers” have converging points of view, the assumption being that, in this case, a common area of assessment conceptions would gather students and teachers.

This article aims also to study, in which classrooms the students perceive exactly teachers' conceptions. The results show different postures among teachers and students, these profiles are very similar. Otherwise, there are very few classrooms within which the actors, “students and teachers” have convergent points of view.

- Jorro, A., Mercier-Brunel, Y. (2011): Les gestes évaluatifs de l'enseignant dans une tâche de correction collective. *Mesure et Évaluation en Éducation*, 34(3), 27-50.

The collective correction is a common practice in the everyday life of a class of years 4-6. So it is rarely called into question by the teachers who have appeal there, and responds most of the time to stereotypes which determine the progress. However, the corrective sessions are important with respect to the quality of the regulation of learning. The teacher practice and, in particular, evaluative gestures determine the entry of students in such processes. While many studies have highlighted the importance of feedback processes in learning, it remains that the implementation of such feedback is still poorly documented. This research aims to identify and characterize the actions of the evaluative situation remedial teacher and their impact on the activity of learners.

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

This paper is already presented in question 8. Here is our translation of the abstract.

The feedback is often considered as a key element to strengthen the motivation and support students' success. In their review of literature, the authors try to highlight the issue. First, the authors recall the different meanings given to feedback and they situate with regard to close terms. Secondly, the authors present their main categorizations before proposing a typology of feedback at school. In a third part, the authors focus on a particular category “attributional feedback”. In conclusion and at this stage of knowledge, the authors ask for some precautions regarding the use of feedback related to attentional resources they may require.

- Talbot, L. (2011): Les pratiques d'évaluation orale des enseignants du primaire et du secondaire. *Mesure et Évaluation en Éducation*, 34(3), 79-112.

This paper deals with the study of the practices of teachers in relation with student learning strategies within the classroom. We will focus in particular on how individual oral assessment is carried out. Our aim is to describe, understand and explain the work of nine teachers, five in primary and the four others in secondary education. It was shown that their verbal assessment habits can be characterized by both intra and between individuals as well as by a certain number of stabilities. Through the use of vid-

eo-taped sequences and statistical processing, we can make sense of these variations and stabilities by studying the correlation between the level of teaching, the level of education and the gender of the student population.

- 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*, 137-162.

The research conducted in the multidisciplinary *Lingot* project concerns the development of Computer-Based Learning Environments (C.B.L.E.) to help teachers take into account the diversity of students' cognitive profiles. It aims to provide teachers with tools to manage the heterogeneity of learning needs in elementary algebra for students at the end of compulsory schooling. First, we show how the *Lingot* project draws on cognitive and anthropological approaches. Second, we outline how didactical models of diagnostic assessment have been developed in the *Lingot* project through the use of automated computer processes. Finally, we present some of the elements of learning routes adapted to the diagnosis and management of interactions differentiated according to pupils' responses.

- Mercier-Brunel, Y., & Jorro, A. (2009): La parole évaluative de l'enseignant. *Les Dossiers des Sciences de l'Éducation*, 22, 9-24.

The collective correction is a common practice in the everyday life of a class of years 4-6. So it is rarely called into question by the teachers who have appeal there, and responds most of the time to stereotypes which determine the progress. However, the corrective sessions are important with respect to the quality of the regulation of learning. The teacher practice and, in particular evaluative gestures, determine the entry of students in such processes. While many studies have highlighted the importance of feedback processes in learning, it remains that the implementation of such feedback is still poorly documented. This research aims to identify and characterize the actions of the evaluative situation remedial teacher action of evaluative situation remedial teacher and their impact on the activity of learners.

- Talbot, L. (2009): *L'évaluation formative: comment évaluer pour remédier aux difficultés d'apprentissage*. Paris: Armand Colin, p. 191.

This is the translation of the French abstract made by us to introduce the book content.

Formative assessment, that is a continuous process of assessment allows the teacher and the students to know at any moment where the constructions of knowledge are, to identify the potential difficulties and consequently to regulate teaching or learning practices. Respectful of behavior differences and students' performances, formative assessment leads to individualize and offer a variety of learning paths for each student. Formative assessment is therefore a valuable help to the most fragile students.

The purpose of the book is to show possible links between assessment practices, learning and remediation. It is intended for trainers, teachers, researchers and students concerned with issues related to learning difficulties and teaching or training practices.

- Bardi, A.M., Mégard, M. (2009): L'évaluation des élèves en France, à un moment charnière de leur histoire ? *Mesure et Évaluation en Éducation*, 32(3), pp. 125-152.

French national education is highly centralized: programs, examination timetables and topics are national, and up to now each school's independence was close to nonexistent. Those equal objectives and means were supposed to result in and guarantee equal results. Therefore, schools aren't subjected to evaluations, and pupils' level and acquired skills were until recently only methodically assessed upon their entering key levels, as a checkup. Throughout "college" (11-15 years old) and "lycée" (15-18 years old, test results are still only transcribed as a numerical grade, and in "école primaire" (6-11 years old), skills assessment booklets are used that do not provide very specific information. However since 2005, redefining texts have been aiming at an altogether new culture of evaluating for the French schooling system: - a financial law requires that the Ministry of Education provide an annual statement on pupils' acquirements, - the law "for the future of school" demands that compulsory schooling provide all pupils with the means they need for reaching basic knowledge and skills. – and European scheme decree specific goals regarding language skills. All the conditions allowing an evolution of (our) current evaluating methods are gathered, whether French education with know to seize that opportunity remains to be seen.

- Antibi, A. (2007): *Les notes: la fin du cauchemar ou en finir avec la constante Macabre*. Math'adore – Nathan, p. 158.

This is the translation of the French summary made by us to introduce the book contents.

The book "The constant macabre", published in 2003, has aroused great interest not only in France but also at the international level. In this book, a very serious deficiency in the assessment system of students, whose teachers are not responsible, is analyzed and denounced: under the pressure of society, teachers unconsciously feel forced to put a certain percentage of bad marks to be credible. A very important movement was born, in order to eliminate the dysfunction. This action has the support of all associations and all the teachers' syndicates, students, students' parents, and many politicians. This second book (our reference) is accessible to everyone, very clear, easy and enjoyable to read, follows the previous book. An "assessment by trust agreement" ("un système d'évaluation par contrat") already practiced by thousands of teachers is presented. In this system, marks truly reflect the value of the student. The results are clear: the "constant macabre" is deleted, and students, more confident, work much more. An overlook on what happens abroad permits to notice that France is isolated in the field of evaluation. A very important survey of teachers, students and parents of students shows clearly very encouraging responses for our educational system. These reactions give hope that some change can take place very soon.

- Dubus, A. (2006): *La notation des élèves: comment utiliser la docimologie pour une évaluation raisonnable*. Paris: Armand Colin, p. 270.

"Who says "marks" says "problem of marking". Whatever can be said or thought on the subject, it remains that marking practices persist imperturbably at elementary school, middle school, high school and university ... Based on that, this book has therefore a pragmatic aim: how we, as teachers, can manage these tedious perdurable practices of evaluation? Since, whatever the reason behind, we continue to give marks to our students' work, how could we do it reasonably? The author focuses on the links that can be made between the contribution of research and the analysis of real cases. Based on all the lessons learned from his teaching experience and on his work on the subject, he provides us with an instrument easy to mobilize for analyzing the practice and the everyday situations of marks attributions for individual or collective use. This book is addressed to teachers, trainers and educational leaders at all levels, as well as to students who are preparing for these professions, but also to anyone who, involved in the educational process, wants to understand better the issue of notation.

Additional information from the authors

Technology is a school subject in France. There's not so much research about evaluation in France. Moreover, the research about evaluation is mostly done by researchers in education, not researchers in science education.

2.6 Germany: A synthesis of the research on assessment of IBE

Sascha Bernholt, Silke Rönnebeck¹⁰, Mathias Ropohl, Olaf Köller, & Ilka Parchmann

2.6.1 Foreword

Based on the non-satisfying results of German students in TIMSS 1995 and PISA 2000 and the huge public debate that arose from that, policy makers took a number of different measures. Programs to improve instructional quality were created (e.g. SINUS – Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts/Increasing the efficiency of mathematics and science instruction), funding of schools was in part increased, and a reform of the education system itself was initiated to address the deficits of German science and mathematics education revealed by PISA. As one consequence, national education standards for different grade levels were developed and implemented in several subjects including (for the end of grade 10) biology, chemistry, physics, and mathematics. These educational standards are obligatory for all federal states and define competences and skills that students should have acquired by the end of primary, resp. lower secondary education. It will be monitored in regular intervals whether students successfully reach them. The standards explicitly mention inquiry-oriented competencies like e.g. the planning, conducting and analysing of experiments, communication or reasoning and argumentation. To which extent such methods are applied in practice to develop the required competencies cannot be stated yet. However, influence of (external) summative assessments on science and mathematics education in Germany has undeniably increased during the last 15 years.

With respect to formative assessment, a review of the empirical German literature on this issue, conducted in 2005, came to the conclusion that there is not very much German research on effects of formative assessment on educational outcomes (Köller, 2005). Although several approaches to formative assessment exist, they have not been sufficiently evaluated. The literature discussed in the following will thus not focus narrowly on empirical studies of formative assessment of inquiry-based competences but present a broader picture of literature on IBE and assessment in science and mathematics education. No information was found concerning technology education which probably is due to the fact that technology is not a regular, compulsory subject in most federal states (or is not taught at all).

2.6.2 Answers to the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

Traditionally, summative assessment and grading play an important role in the German educational system. In almost every subject there are several written assessments during the school year that have a significant impact on the final grades – and subsequent tracking. The teacher is often the only evaluation authority whose criteria, however, are mostly not made explicit (Winter, 2007). Large-scale external assessments, on the other hand, do not have a long tradition in Germany. One reason for this is seen

¹⁰ Corresponding author (email: roennebeck@ipn.uni-kiel.de)

in the paradigm of *Bildung* that for a long time formed the core of the German educational system (Oelkers, 2001; Neumann, Fischer, & Kauertz, 2010). The concept of *Bildung* goes back to the German scholar, statesman and co-founder of the University of Berlin, Wilhelm von Humboldt. *Bildung* is regarded as a personal attitude towards learning (Oelkers, 2002) that should support students in developing a critical awareness thus enabling them to evaluate situations, make independent decisions and act responsibly (Universität Mainz, 2013). Within the paradigm of *Bildung*, knowledge is always subjectively bound. Consequently, large-scale assessments were not considered meaningful.

However, the growing need for global competitiveness led to Germany's participation in TIMSS 1995 and later on in PISA. The results of TIMSS 1995 and PISA 2000 came as a negative surprise to all stakeholders in the educational system, leading to the aforementioned debate and consequences. The core of the resulting reform was a shift from a system in which educational policy is implemented by exerting control over the input to one in which educational policy is implemented by defining and controlling the output. This led to the development and implementation of education standards that had been agreed on and are therefore obligatory in all states; even though education is subject to federal state authority. Thus each federal state has its own school system, including the definition of types of schools, curricula and assessment regulations. From now on, the attainment of the education standards will be monitored in regular intervals by national large-scale assessments ("Ländervergleiche"). The aim, however, is to survey the educational system and not to yield information at the student level. Therefore, assessment in class, set up by the teachers during the courses and the teachers or the states for the 'Abitur' will additionally still be used for grading and transition.

Despite the long tradition of grading, in the 1980s an attempt was made to replace grades by verbal reports in primary school (at least for grades one and two). The aims were, among others, to increase students' individual support and base assessment more on individual progress than on social comparisons. However, empirical evidence showed that the reform did not have the intended outcomes and children did not profit as much from them as it had been predicted (e.g. Benner & Ramseger, 1985; Wagner & Valtin, 2003).

Formative assessment in Germany has its most important historical root in the so-called "Reformpädagogik" (engl. reform pedagogy or alternative education) that has a long tradition among German philosophers and educational reformers (Köller, 2005). Fundamental ideas include self-assessment, autonomy, and self-regulation. Moreover, approaches to and the handling of feedback play an important role within this movement and have ever since been a focus of research on formative assessment in Germany (see later sections).

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?

No publications were found dealing with the extent of the use of different formats of structured formative assessment. However, in the review by Köller (2005) formats typically employed in Germany are mentioned. Those formats include:

- Diagnostic forms (“Diagnosebögen”) that provide detailed information about individual learning success. They were mostly used at comprehensive schools but abandoned in the 1980s because they were perceived to be too time-consuming by teachers and school administrators.
- Learning reports (“Lernberichte”, also Lübke, 1996) that are alternative forms of summative assessment and combine information about social and cognitive learning outcomes.
- Learning diaries (“Lerntagebücher”, also Herrmann & Höfer, 1999) that provide opportunities for students to reflect on their own learning processes thus serving as a tool for autonomous and self-regulated learning.
- Weekly working plans (“Wochenarbeitspläne”) that allow students to monitor whether they have reached their goals for that week.
- Portfolios that are considered to be particularly useful in cooperative learning environments (e.g. Gläser-Zikuda, 2007).

Another aspect that has gained importance over the last decade is the development and implementation of scoring rubrics to increase the transparency of the evaluation process (Schrempf, 2002; Lissmann, 2007). Within the project *Chemie im Kontext* different instruments were developed and implemented that serve formative assessment purposes. So-called accompanying reflection sheets (“Lernbegleitbögen”) were used before, during and after instructional units. They consisted of questions related to everyday phenomena that the students repeatedly had to answer. Students got their answers back allowing them to comment on their ideas, correct them if necessary and thus follow their own learning progress (Demuth, Gräsel, Parchmann, & Ralle, 2008). Other instruments developed within the project deal with the use of student experiments as an assessment tool (di Fuccia, 2007). Instruments that were used required students e.g. to fill-in-the-blanks in existing experiment descriptions, to develop experiment instructions by themselves, to conduct experiments without instructions and to predict observations. It was found that student experiments are principally suited for assessment purposes. The evaluation of student experiments provide teachers with valuable information about their students learning; feedback of the evaluation results to the students, however, works only partially (di Fuccia, 2007).

Assessment, evaluation and feedback are also addressed within the SINUS-project. SINUS is a nationwide program that aims at the improvement of science and mathematics instruction in Germany (for an overview about SINUS in English see Prenzel, Stadler, Friedrich, Knickmeier, & Ostermeier, 2009). To reach this goal, SINUS focuses on the professional development of teachers who work collaboratively on so-called “modules” that address problem areas of science and mathematics education in Ger-

many. Characteristics of “good tasks” – both for learning and for assessment – play an important within SINUS. At the lower secondary level, the module that was worked by far the most on, dealt with this issue. Another module specifically addresses aspects of assessment, i.e. surveying and providing feedback on competency increases. An overview about instruments that have been developed within this context is e.g. given in Vollstädt, W. (2005).

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

There were not many research results found to answer this question. A teacher characteristic, however, that positively influences the use of formative assessment, especially peer- and self-assessment, is a high diagnostic competence (Klieme et al., 2010). Moreover, we would hypothesize that the factors mentioned in the international literature also apply for Germany. Teachers need support not only in the recognition of opportunities for formative assessment, the development and use of formative assessment instruments, and the provision of developmental feedback, but also in changing their pedagogical beliefs about assessment. They have to perceive formative assessment as a natural part of their instruction and everyday teaching and not as something they are supposed to do in addition.

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results used formatively and/or vice versa?

The development of education standards in Germany was based on an expertise produced by a high-level group of educational experts (Klieme et al., 2003). The authors explicitly stress that the standards can give hints for individual diagnosis and support. However, due to methodological constraints, assessments that aim at monitoring the educational system and evaluating schools usually do not allow for individual diagnosis. Grading – that has traditionally a summative function – is regarded as the most important form of formative assessment in Germany (Klieme et al., 2010).

Within the research project “Conditions and Consequences of Classroom Assessment (Co²CA)” Klieme and colleagues (2010) investigated the question whether classroom based assessments and standard-based tests measure psychometrically independent constructs. In a study with more than 1500 students from 9th grade mathematics classrooms they found that it is in principle possible to represent thematically focused (classroom-based) and broader (standards-based) test items in common competency models. On the basis of item analyses, however, also specific competence models for sub-dimensions could be developed. The empirical data better fit to a multidimensional than to a global one-dimensional model. The authors thus concluded that thematically focused assessments – as needed for formative assessment – led to additional and specific information that could be provided by summative assessments.

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

Again, there are not many research results available concerning this topic. As described under question three, teachers need to feel diagnostically competent to make use of formative assessment (Klieme, 2010). Within the above-mentioned Co²CA-project it is planned to perform additional analyses concerning the relationship between teacher characteristics, instructional and assessment practices, and their influence on the use of formative assessment (Bürgermeister et al., 2011). Moreover, we would again hypothesize that the results found in the international literature like deeply held pedagogical beliefs and time constraints are also valid for Germany (like e.g. discussed in the Swiss national report (Smit, 2009)). Another aspect is brought in by Maier (2010) who argues that researchers and educators in Germany do not make enough use of the international discussion about formative assessment (Maier, 2010).

Winter (2007) eventually discusses the dilemma between alternative assessment methods that aim at the contemplation of learning (like e.g. portfolios or learning diaries) and student evaluation. An open discourse about learning paths, conditions, goals, and output is opposed to a concept of grading. If students fear that their expressed opinions are graded they will possibly withhold them. On the other hands, students might be demotivated if they put much effort e.g. into a portfolio that virtually does not influence their grades at all. He considers different forms of feedback as appropriate assessment formats for portfolios and learning diaries. These are personal resonance (“What I liked about your work”), finding qualities (“What I think you did really good”), revising (“Where and how you could improve), and relating the work to objective scoring criteria (e.g. rubrics)

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

In this context, one could again mention the implementation of the national education standards. They include inquiry competences and thus require e.g. the development of competence models for IBSE for monitoring purposes. Mayer, Grube, & Möller (2009) tested a predicted model of scientific inquiry skills with more than 1500 students in grades 5-10. At the same time, a test instrument with 24 open response items was developed and validated. The items aimed to assess different subcompetences of scientific inquiry (identifying questions, formulating hypotheses, planning investigations, analysing data, and drawing conclusions). The analysis led to a 4-dimensional model with a considerable amount of common variance. The items allowed for differentiating subcompetences and showed good discriminating validity with respect to biological content knowledge.

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding

of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

Inquiry competences are part of the national educational standards in Germany and will thus be assessed on a regular basis in nationwide large-scale assessments. However, the first assessment of science and mathematics took place in 2012 and the results have not yet been published.

Moreover, there are several small to medium scale studies in Germany dealing with the assessment of inquiry competences in biology, chemistry, and physics that will be shortly described in the following.

In the field of biology education, Hammann and colleagues investigated the assessment of students' experimentation skills. They understand experimentation as a problem-solving process consisting of three subdimensions, namely searching in the hypotheses space (1), testing of hypotheses (2), and analyzing evidence (3) (Hammann, Phan, & Bayrhuber, 2007).

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.

The authors operationalized the subdimensions of experimentation as multiple choice items and analyzed the reliability of the scales as well as intercorrelations between the three subdimensions and their respective relations to prior biological content knowledge. The study included 1006 students in grades 5 and 6. The scales showed satisfying reliabilities. However, confirmatory factor analysis revealed that the test measured only two dimensions, namely one depending on prior knowledge (searching in the hypotheses space and analyzing evidence) and one depending upon methodological knowledge about the aims and processes of experimentation (testing of hypotheses). The subdimensions showed moderate to high intercorrelations. The correlation between the subdimensions and the biological content knowledge test are lower than the intercorrelations which might be partly due to the lower reliability of the knowledge test.

Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008): Assessing pupils' skills in experimentation. Journal of Biological Education, 42(2), 66-72.

This study used the same MC test as the study above. It compared different forms of assessment of pupils' skills in experimentation: 1. the MC test with an open response test to analyse pupils' skills in planning two-factor experiments (323 grade 6 students), and 2. the MC test with a practical test on seed germination (24 grade 5 students). With respect to the first study, a low correlation between the MC and open response test was found. Many open responses lacked the ideas of comparison or experimental controls. The authors concluded that the two test formats might be measuring different

constructs. The results of the second study showed that the actual performance in a 'real' experiment was positively correlated with the performance in the multiple choice test. The low correlation coefficient of 0.33, however, indicated that successful performance in a multiple choice test was not a very strong predictor for successful performance in the more open situation of planning an actual experiment.

Also located in the field of biology education is a study by Urhahne, Kremer, & Mayer (2008) describing the development and first steps towards the validation of a questionnaire assessing the nature of science.

Urhahne, D., Kremer, K., & Mayer, J. (2008): Welches Verständnis haben Jugendliche von der Natur der Naturwissenschaften? Entwicklung und erste Schritte zur Validierung eines Fragebogens [What is adolescent's understanding of the nature of science? Development and first step of validation of a questionnaire]. Unterrichtswissenschaft, 36(1), 71-93.

The questionnaire consists of 111 items from established instruments assessing eight key dimensions of NOS belonging to two domains: ideas about scientific knowledge and ideas about scientific methods. It was tested with 272 students (grades 6-10). As a result, seven dimensions could be empirically verified. The instrument seems to be valid but the scales show rather low reliabilities. This might be due to a rather low significance of NOS for German classroom practice. The understanding of NOS often seems to be unsatisfactory.

In the field of chemistry education, Walpuski & Schulz analyzed the relation between inquiry competences and experimentation by using video analysis (Walpuski & Schulz, 2011).

Walpuski, Maik, Schulz, Alexandra (2011): Erkenntnisgewinnung durch Experimente [Inquiry through experiments]. chimica et ceterae artes rerum naturae didacticae, 37 (104), 6-27.

In a first study, a video analysis of 336 seventh grade students working on open, experimental inquiry tasks was performed. The results showed gains in content knowledge but problems especially in the planning and conducting of experiments thus stressing the importance of appropriate feedback. In a second study, 264 grade 10 students took part. Their chemistry lessons were videotaped and the students' content knowledge and inquiry competence were assessed (the latter by the NAW test). The results showed that experimentation made up for more than 90% of instructional time but in more than half of the cases, the experiments were not embedded in theory. Moreover, the students hardly worked in a hypotheses-oriented way and they only tried to verify instead of falsifying their hypotheses.

Björkman, J., Labetzki, T. & Tiemann, R. (2012): Ein Instrument zur Videoanalyse von Scientific Inquiry [An instrument for the video-based analysis of scientific inquiry]. In S. Bernholt (Hg.), Konzepte fachdidaktischer Strukturierung für den Unterricht. GDGP, Jahrestagung in Oldenburg 2010. Münster: LIT, 304-306.

The aim of this study was to develop a reliable Instrument for the video-based analysis of scientific inquiry in chemistry instruction by constructing a coding guide to identify instructional patterns. The study was located at the teacher level. It defined variables related to three phases of inquiry: 1. Research question and hypotheses, 2. Planning and conducting an investigation, and 3. Evaluation, interpretation, and reflexion. The variables showed a moderate to very good coder reliability. A prevalence of the phases one and two could be observed. This might be due to the fact that teachers usually do not let students generate their own questions and planning their own experiments and thus needed considerably more time for these two steps.

Another publication deals with the assessment of experimentation in physics education (Schreiber, Theyßen, & Schecker, 2009).

Schreiber, N., Theyßen, H., Schecker, H. (2009): Experimentelle Kompetenz messen?! [Assessing experimentation competence?!]. *Physik und Didaktik in Schule und Hochschule*, 3(8), 92-101.

The aim of the study was to compare three test formats – paper-and-pencil, real experiment, and a simulation kit – to assess experimentation competence. It was inspired by an international study by Shavelson et al. (1999). This study compared different assessment formats and found that observation, student notebooks, and simulation kits are exchangeable whereas paper-and-pencil and performance experiments are not. In the German study, the authors use a model of experimentation consisting of the three subdimensions planning an experiment, conducting an experiment, and analysing/evaluating results. The paper describes the model as well as the design of the instrument and the planned investigation (no empirical results are presented).

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

Within the project Co²CA several publications related to the use and effects of feedback in mathematics instruction were published that will be presented in the following.

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

The study aimed at answering the following research questions: 1. Are classroom based assessments and standard-based tests measuring psychometrically independent constructs?, 2. Which assessment practice dominates in mathematics classrooms? Are the practices related to teacher characteristics?, 3. Which forms of assessment are related to high levels of achievement and test motivation?, and 4. How are different forms of feedback (criterion-referenced, social comparative and process-referenced) related to satisfaction and attribution?

To answer these questions, 138 mathematics items were developed in two content areas (Pythagoras theorem and linear equations) and two competence dimensions (modeling and technical competence). In addition, 38 items of the standards-based mathematics assessment were used. The items were administered to 1560 students in the 9th grade. Six months after the assessment, a sample of 167 students received feedback about their performance. Feedback was provided in three different forms: criterion-referenced, process-referenced or social comparative. The results show that in the classrooms investigated in this study verbal feedback was dominant combined with different forms of teacher- and grade-centered assessment. Only 10% of the teachers explicitly defined assessment criteria before tests or exams and teacher/grade-centered assessment lead to low levels of test motivation. On the contrary, formative assessment correlated to higher levels of motivation. Especially, criterion-referenced feedback increased the satisfaction with the feedback results and the tendency for internal attribution (compared to social-comparative feedback); no effect was observed for process-referenced feedback. Self- and peer-assessment were generally rare (even less common were portfolios or learning diaries) but comparatively more common if teachers (especially female teachers) self-reported high diagnostic competences. High diagnostic teacher competences are moreover related to high student achievement on the test items.

Rakoczy, Katrin, Klieme, Eckhard, Bürgermeister, Anika, & Harks, Birgit (2008): The Interplay Between Student Evaluation and Instruction - Grading and Feedback in Mathematics Classrooms. Zeitschrift für Psychologie/Journal of Psychology, 216(2), 111-124.

This study aimed at evaluating the impact of evaluative and informational feedback provided in actual classroom settings on students' motivation and achievement. It was conducted with 240 students and used video analysis, interviews, and an achievement test focusing on the conceptual understanding of the Pythagorean theorem and on simple application tasks. It could be shown that positive evaluative feedback in the classroom is associated with increased intrinsic motivation whereas negative evaluative feedback is not related to motivation. Informational feedback, as observed in actual classroom interactions, seemed to foster motivation via emotional experience and cognitive processing. None of the feedback types examined had a significant impact on students' achievement development.

Rakoczy, Katrin, Harks, Birgit, Klieme, Eckhard, Blum, Werner, Hochweber, Jan (2013): Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. Learning and Instruction, 27, 63-73.

This study aimed to investigate the following research questions: 1. Does process-oriented feedback in mathematics leads to greater interest and higher achievement development compared to social-comparative feedback?, 2. Does students' perception of feedback with regard to usefulness and competence support mediate these effects?, and 3. Is the impact of feedback moderated by students' mastery approach goal orientation?

146 ninth-grade students participated in the study. Students were assigned to one of two written feedback conditions, process-oriented or social-comparative. Perceived competence support, usefulness, development of interest, and mastery approach goal orientation were assessed by analyzing students' responses to questionnaires. Mathematics achievement was assessed by administering curriculum-embedded mathematics tests either belonging to the content domain of the Pythagorean Theorem or linear equations, or originating from the item pool of the German national educational standards. Feedback was given immediately after a pretest (consisting of scored test plus written feedback plus explanations by the test administrator). Afterwards, students' perceived competence support and usefulness of feedback and interest in a second forthcoming test were assessed followed by a posttest. The authors found no significant total feedback effects on interest and achievement development. There were, however, indirect effects on the development of interest via the perceived competence support and usefulness, and on achievement development via the perceived usefulness. A mastery approach goal orientation mediated the impact of feedback on the perceived usefulness.

Bürgermeister, A., Klimczak, M., Klieme, E., Rakoczy, E., Blum, W., Leiss, D., Harks, B., & Besser, M. (2011): Leistungsbeurteilung im Mathematikunterricht [Performance assessment in mathematics instruction]. Schulpädagogik – heute, 22, 1-18.

This paper gives an overview about different phases and sub-studies of the project "Conditions and Consequences of Classroom Assessment (Co2CA)". In particular, results of a laboratory study and the design of a classroom study are presented.

The research questions for the laboratory study were as follows: 1. How does the type of feedback influences achievement, motivation, and metacognitive variables?, 2. Which variables mediate the influence of feedback on achievement and motivation?, 3. Is influence of feedback dependent on learner characteristics?, and 4. Are the results influenced by the proximity of the test to the curriculum?

As in the other Co²CA publications, the mathematical content areas Pythagorean theorem and linear equations and the two mathematical competences modeling and technical competence were investigated. Modeling competence is assessed by realistic tasks in which students are required to construct and work with a mathematical model and interpret and evaluate the result. 329 ninth grade students participated in the study. The assessment design principally followed Rakoczy et al. (2013). However, in this study five feedback types were differentiated: social comparative, process-oriented, criterion-differentiated, criterion-non-differentiated, and no feedback at all. The results show that process-oriented feedback was found to have a more positive impact on motivation and development of (both) mathematical competences as social-comparative feedback. Individual feedback including weaknesses, strengths, and strategies for improvement are more helpful for students in adapting their learning processes than grades. Process-oriented feedback is moreover perceived by students to be more supportive and useful than social-comparative feedback. Results with respect to the influence of the proximity of the test to the curriculum are not yet available.

The presented classroom study aims at answering the following questions: 1. Does formative assessment has an impact on student achievement and motivation?, 2. What are effects of teacher characteristics, assessment practice, and instructional practice on achievement and motivation and how are they influencing formative assessment?, 3. Which students profit from formative assessment?, 4. Is the integration of individual written and verbal feedback in the classroom feasible?, and 5. Can formative assessment sustainably support students' self-regulated learning, autonomy, and competence development? This study will be based on 41 teachers and approx. 1200 students. Results are not yet available.

The effects of different forms of feedback in mathematics was also analyzed by Narciss & Huth (2006). In a computer-based learning environment for written subtraction with 50 fourth-grade students, cognitive and motivational effects of bug-related tutoring feedback (BRT) compared to traditional knowledge of result (KR) and knowledge of correct result (KCR) feedback were investigated. BRT feedback indicates the correct response but offers in addition explanations for the correction of errors. Results showed that BRT-feedback is significantly more beneficial for motivation and achievement than KR-KCR-feedback.

Gläser-Zikuda and Lindacher (2007) investigated the effects of the use of portfolios on cognitive and affective measures (however, not in STM but in history classrooms). In an experimental study with approximately 40 students in grade 10, they found a statistically significant and sustainable effect on learning performance (measured by a test focusing on factual and transfer knowledge). The portfolio approach led to high levels of interest, motivation, and acceptance on the side of the students which was attributed to the fact that ownership was given to the students.

A study in chemistry education compared the use of two evaluation instruments, namely concept maps and accompanying reflection sheets, to support students in the (self)-diagnosis of their conceptual understanding (Schanze, Grüß-Niehaus, & Hundertmark, 2011). The authors could show that both instruments can be used to support students' (self) diagnosis. The acceptance of the instruments among the students was generally positive with a significant preference for the concept mapping technique.

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

Especially the results from large-scale assessments, either international (PISA, TIMSS) or national (state comparisons based on education standards) receive a lot of attention from all stakeholder groups in Germany. A tendency can be observed to make educational outcomes more comparable, e.g. by so-called "Vergleichsarbeiten" (engl. comparative tests) at school or federal state level or by discussions about whether the upper secondary school leaving certificate ("Abitur") should be the same in all states

Question 10: Which are the 10 most relevant publications on formative and summative assessment of inquiry in science, technology, and mathematics in your country (please make sure that your selection covers all three subject areas)? If there are no publications especially dealing with IBE, please list the 10 most relevant publications in forma-

tive and summative assessment in general. What are the main results of these publications (for results refer to earlier sections)¹¹?

- Björkman, J., Labetzki, T. & Tiemann, R. (2012). Ein Instrument zur Videoanalyse von Scientific Inquiry [An instrument for the video-based analysis of scientific inquiry]. In S. Bernholt (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht. GDCP, Jahrestagung in Oldenburg 2010* (pp. 304-306). Münster: LIT.
- Bürgermeister, A., Klimczak, M., Klieme, E., Rakoczy, E., Blum, W., Leiss, D., Harks, B., & Besser, M. (2011). Leistungsbeurteilung im Mathematikunterricht [Performance assessment in mathematics instruction]. *Schulpädagogik – heute*, 22, 1-18.
- 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 - Sonderheft*, 8, 33-49.
- Hammann, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42(2), 66-72.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D. & Rakoczy, K. (2010). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht [Assessment and modeling of competences within mathematics instruction]. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (pp. 64-74). *Zeitschrift für Pädagogik* 56. Beiheft. Weinheim, Basel: Beltz.
- Rakoczy, Katrin, Harks, Birgit, Klieme, Eckhard, Blum, Werner, Hochweber, Jan (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63-73.
- Schanze, S., Grüß-Niehaus, T., & Hundertmark, S. (2011). Verstehen sichtbar machen. Instrumente zur Unterstützung der (Selbst-) Diagnose des Konzeptverständnisses [Making understanding visible. Instruments to support (self-) diagnosis of conceptual understanding]. *Unterricht Chemie*, 22(124/125), 68-74.
- Schreiber, N., Theyßen, H., Schecker, H. (2009). Experimentelle Kompetenz messen?! [Assessing experimentation competence?!]. *Physik und Didaktik in Schule und Hochschule*, 3(8), 92-101.
- Urhahne, D., Kremer, K., & Mayer, J. (2008). *Welches Verständnis haben Jugendliche von der Natur der Naturwissenschaften? Entwicklung und erste Schritte zur Validierung eines Fragebogens* [What is adolescent's understanding of the nature of science? Development and first step of validation of a questionnaire]. *Unterrichtswissenschaft*, 36(1), 71-93.
- Walpuski, Maik, Schulz, Alexandra (2011). Erkenntnisgewinnung durch Experimente [Inquiry through experiments]. *chimica et ceterae artes rerum naturae didacticae*, 37(104), 6-27.

¹¹ The English titles are (mostly) translations by the authors of the report.

Moreover, there are some basic publications about inquiry-based education in Germany:

Bell, T. (2007). Entdeckendes und forschendes Lernen [Inquiry learning]. In S. Mikelskis-Seifert & T. Rabe (Hrsg.), *Physikmethodik. Handbuch für die Sekundarstufe I und II*. Berlin: Cornelsen Verlag Scriptor.

In this contribution, the authors describe e.g. problems students have with different aspects of inquiry e.g. identifying questions, identifying variables, or formulating testable hypotheses. Assumptions are often investigated unsystematically and students tend to rather confirm a hypothesis than reject it (thereby often reinterpreting data). Several computer-based learning environments e.g. Knowledge Forum, WISE, Viten, and Co-Lab are named.

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen [Inquiry as scientific problem-solving]. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (pp. 177-186). Berlin, Heidelberg, New York: Springer.

Reitinger, J. (2013). *Forschendes Lernen. Theorie, Evaluation und Praxis in naturwissenschaftlichen Lernarrangements [Inquiry-based learning. Theory, evaluation and practice in scientific learning environments]*. Theorie und Praxis der Schulpädagogik, Bd 12. Immenhausen, Hess: Prolog-Verlag.

Schmidkunz, H. & Lindemann, H. (1992). *Das forschend-entwickelnde Unterrichtsverfahren. Problemlösen im naturwissenschaftlichen Unterricht [The concept of inquiry-based teaching. Problemsolving in science instruction]*. Hohenwarsleben: Westarp Wissenschaften.

Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht - Ergebnisse einer Videostudie [Experimentation in physics instruction – Results of a video study]. *Zeitschrift für die Didaktik der Naturwissenschaften*, 10, 51-69.

In this study, video analysis was performed to analyze and characterize experimentation practice in German physics classrooms. A coding scheme for practical activities developed for an analysis of video-taped physics lessons is developed and presented. Results show that on average 64% of the total lesson time is influenced by practical activities, although actually carrying out experiments is observed in just 28% of the total lesson time.

References

Benner, D. & Ramseger, J. (1985). Zwischen Zifferzensur und pädagogischem Entwicklungsbericht. Zeugnisse ohne Noten in der Grundschule [Between numerical grades and pedagogical reports on development. School reports without grades]. *Zeitschrift für Pädagogik*, 31, 151-174.

Demuth, R., Gräsel, C., Parchmann, I., & Ralle, B. (2008). *Chemie im Kontext. Von der Innovation zur nachhaltigen Verbreitung eines Unterrichtskonzeptes [Chemistry in Context. From the innovation to the sustainable dissemination of an instructional concept]*. Münster: Waxmann.

- Di Fuccia, D.S. (2007). *Schülerexperimente als Instrument der Leistungsbeurteilung* [Student experiments as an instrument of performance evaluation]. Dissertation. Berlin: uni-edition.
- Gläser-Zikuda, M. (2007). *Lernprozesse dokumentieren, reflektieren und beurteilen: Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* [Document, reflect on, and evaluate learning processes: Learning diary and portfolio in educational research and praxis]. Klinkhardt: Bad Heilbrunn.
- Gläser-Zikuda, M. & Lindacher, T. (2007). Portfolioarbeit im Unterricht – praktische Umsetzung und empirische Überprüfung [Portfolios in instruction – practical implementation and empirical validation]. In: Gläser-Zikuda, M. (Hrsg). *Lernprozesse dokumentieren, reflektieren und beurteilen: Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* (pp. 189-204). Klinkhardt: Bad Heilbrunn.
- Hermann, J. & Höfer, C. (1999). *Evaluation in der Schule – Unterrichtsevaluation* [Evaluation in schools – evaluation of instruction]. Bertelsmann-Stiftung: Gütersloh.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise* [About the development of national education standards. An expertise]. Bonn: BMBF.
- Kölller, O. (2005): Formative Assessment in Classrooms: A Review of the Empirical German Literature. In OECD (Ed.), *Formative Assessment. Improving Learning in Secondary Classrooms* (pp. 266-280). Paris: OECD.
- Lissmann, U. (2007): Beurteilungsraster und Portfolios [Scoring rubrics and portfolios]. In: Gläser-Zikuda, M. (Hrsg), *Lernprozesse dokumentieren, reflektieren und beurteilen: Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* (pp. 87-108). Klinkhardt: Bad Heilbrunn.
- Lübke, S.-I. (1996). *Schule ohne Noten* [School without grades]. Leske + Budrich: Opladen.
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? [Formative assessment – a promising concept to reform instruction and assessment?] *Zeitschrift für Erziehungswissenschaft*, 13(2), 293-308.
- Mayer, J., Grube, C. & Möller, A. (2009). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung [A competence model of scientific inquiry skills]. *Lehr- und Lernforschung in der Biologiedidaktik*, 3, 63-80.
- Narciss, S. & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction* 16(4), 310-322.
- Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to Educational Standards: The Impact of Large-Scale Assessments on Science Education in Germany. *International Journal of Science and Mathematics Education*, 8, 545-563.
- Oelkers, J. (2001). Wo bleibt das humanistische Bildungsideal? *Universitas. Orientierung in der Wissenswelt. Deutsche Ausgabe*, 56(7), 700-707.
- Oelkers, J. (2002). Und wo, bitte, bleibt Humboldt? *Die Zeit*, 27, 27.06.2002, 36.
- Prenzel, M., Stadler, M., Friedrich, A., Knickmeier, K., & Ostermeier, C. (2009). *Increasing the efficiency of mathematics and science instruction (SINUS) – a*

large-scale teacher professional development programme in Germany. Kiel: IPN.

- Schrempf, R. M. (2002). Rubrics. Ein Instrument zur Qualitätsentwicklung und Qualitätssicherung in Unterricht und Schule [Rubrics. An instrument for quality development and assurance in instruction and schools]. *Pädagogik*, 54 (9), 40-43.
- Shavelson, R. J., Ruiz-Primo, M. A., Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36(1) 61-71.
- Universität Mainz (2013). *Das humanistische Bildungsideal im Diskurs*. Retrieved from: http://www.geschichte.uni-mainz.de/neueste_geschichte/1046.php (15.07.2013).
- Vollstädt, W. (2005). Leistungen ermitteln, bewerten und rückmelden. Qualitätsinitiative SINUS. Weiterentwicklung des Unterrichts in Mathematik und den naturwissenschaftlichen Fächern [Assessment, evaluation and feedback of performance. Quality initiative SINUS. Development of instruction in mathematics and science]. *Materialien zur Schulentwicklung*; 39. Frankfurt am Main: Amt für Lehrerbildung.
- Wagner, C. & Valtin, R. (2003). Noten oder Verbalbeurteilungen? Die Wirkung unterschiedlicher Bewertungsformen auf die schulische Entwicklung von Grundschulkindern [Grades or verbal reports? The effects of different forms of evaluation on the educational development of primary school children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 35, 27-36.
- Winter, F. (2007): Fragen der Leistungsbewertung beim Lerntagebuch und Portfolio [Questions concerning the evaluation of learning diaries and portfolios]. In: Gläser-Zikuda, M. (Hrsg.). *Lernprozesse dokumentieren, reflektieren und beurteilen: Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis* (pp. 109-129). Klinkhardt: Bad Heilbrunn.

2.7 Switzerland: Synthesis of Swiss Research Results on Formative Assessment

Regula Grob¹², Monika Holmeier, & Peter Labudde¹³

2.7.1 Summary

In Switzerland, competency models and new ways of assessment have been under discussion a lot in the recent years. New comprehensive curricula, valid for all cantons ("states") in each of the linguistic regions, have been elaborated. In the course of this, educational standards for different age classes have been defined in several subjects, including science and maths. The demand of the economy to teach students more complex skills and competences has accounted to the efforts to further promote formative assessment in IBSE.

Despite this progress, there is still very little basic research on formative assessment. The research groups who work in the field mostly concentrate on assessment tools and teaching materials. Quite a range of approaches has been described and published in the recent years (will be presented in questions 7 and 8). Almost all of this research is on maths. It should be stressed that there is a long tradition of summative assessment and grading in Switzerland; so the description of the research results does not reflect the average daily teaching practice.

Technics is not taught as a subject at the level of compulsory school; topics similar to the ones formulated in the German "standards in technics for the outcomes of the medium level of compulsory school" ("*Bildungsstandard Technik für den mittleren Schulabschluss*") are integrated in classes such as science, Technisches Gestalten ("*technical design*"; a subject specific for Switzerland where handicrafts and related disciplines are taught) or maths. This might be part of the reason why there is almost no research on technics.

2.7.2 Answers to the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

In Switzerland, there is a long tradition of summative assessment and grading. Typically, there are about 2 -5 written exams per semester in each of the STM subjects, on which the final grades of the school report is based.

Switzerland has no culture of high-stake large-scale assessment. The country does participate in the PISA assessments and has taken part in one TIMMS evaluation (1995). Within the country, there are punctual large-scale assessments (in most of the cantons; Switzerland is a federal republic, and each of the 26 cantons ("states") has its own school system). The main purpose of these assessments is to survey the educational system but not to yield a direct impact on individual students. In addition to these large-scale assessments which have been developed in the recent years, discussions to coordinate summative assessments at the level of form groups within individual

¹² Corresponding author (e-mail: regula.grob@fhnw.ch)

¹³ Corresponding author (e-mail: peter.labudde@fhnw.ch)

schools or even between different schools (e.g. the question if or not the final exam at the end of year 12 should be the same) become more and more prominent.

Grades are generally accepted and thought to be necessary in Switzerland, particularly among parents and students (e.g. Dzelili, 2009). This is, according to the same author, based on the lack of reflection over intent and purpose of grades. The typical flaws of summative assessments are well-known among teachers and researchers (e.g. Frey & Frey-Eiling, 2004); nevertheless, efforts to reduce the importance of traditional tests and grading are difficult to communicate and their political acceptance is not easily achieved (Fischer, 2009).

According to Vögeli-Mantovani, 1999, formative assessment has been an issue in Switzerland since the early eighties and has been discussed, conceptionally consolidated, and tested particularly at primary school level. The fundamental ideas include self-assessment and student's responsibility for their learning. Most states have worked out guidelines on assessment at compulsory school level. These mandatory guidelines include considerations upon different formats and purposes of assessment and explicitly mention formative assessment as frequent, short, straight-forward actions to support student's learning and motivation. There are assumptions, but no systematic surveys on the formative practice. The above-mentioned author also points out the difficulties related to the federal system in Switzerland: the culture of assessment still varies - between different states as well as between different levels of education (Vögeli-Mantovani, 2009). Legal standards allow great individuality even at the level of school units (Husfeld, 2009) and at the level of individual classes (Kronig, 2009).

Currently, one single new curriculum valid for all states of one linguistic region, e.g. the so called Lehrplan 21 ("*curriculum 21*") for the German speaking part of Switzerland, is being developed and will be implemented from 2015/16. In this process, the concept of competency models has become prominent.

Newly developed school books (e.g. Schweizer Zahlenbuch; a maths book series) base on the same fundamental idea and include assistance for teachers in terms of different formats and methods of assessment. There is a common understanding among scientists that assessment and planning of lessons must be oriented at educational objectives and competences, less so among teachers and politicians (Rothenbacher, 2010).

The situation in assessment of STM does not differ a lot from the preceding general remarks about the situation in Switzerland. Talking about maths, Rothenbacher, 2010, states that, particularly for teachers who have grown up with testing declarative knowledge, it is difficult to adapt to the idea of competence orientation and different functions of assessment (summative assessment versus formative assessment). He adds that, particularly in arithmetics, conventional procedures seem to be objectively assessable. And therefore, he claims, they often turn out to be the only competence assessed. This is in consistence with Adamina, 2010, who describes the situation in science. He points out, that assessment is very often reduced to declarative knowledge. As soon as non-declarative knowledge is assessed, he adds, assessment

criteria are not dominated by content- but by formal issues such as length of a talk or layout of slides (Marco Adamina, pers. comm., 2013). He strongly promotes the idea of assessment as a means of support and guidance: helping to develop self-regulated learning, curiosity, interest. He adds that, in order to achieve these aims, assessment must refer to the educational objectives and therefore be integrated in the lesson planning.

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?

It was not possible to find studies covering the question on the extent of structured formative assessment and its formats. Nevertheless, there is data on the attitude to assessment in the format of oral or written feedback given by the teachers (instead of the traditional grades which include no support for further learning).

New ways of assessment such as oral feedback at the end of the semester and learning reports (instead of grades) are highly accepted among teachers. In canton-wide evaluations (canton = "state"), over 90% of teachers rated oral feedback by the teachers positive, and more than 75% thought that learning reports on the progress of their students are important (Vögeli-Mantovani, 1999). Similar results were found on the attitude to self-assessment of the students: 90% of teachers acknowledged their value. Comparably high approval was found among parents and students (Vögeli-Mantovani, 1999).

In the 1980 and 1990, the idea of learning journals (original title: *Reisetagebücher* or *Lerntagebücher*) was developed and promoted in Switzerland. The basic idea was to write down the individual steps towards the solution of complex problems in maths. The students invested more time for writing than in traditional maths lessons. Based on these notes, the teacher would be able to give continuous, personalised advice on the next steps in the student's learning (Ruf & Gallin, 1991). The idea was accepted by many teachers. There is no data on the exact extent of use of these learning journals, though.

Summarizing the practice of formative assessment in Switzerland, it should be stressed that there are several approaches to promote the issue, but not (yet) nationwide. The methods of formative assessment mentioned in this report are more advanced than the average practice at schools. Nevertheless, the foundations for more comprehensive formative assessment are there.

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

Consistent with researchers from other countries, Schwartz & Allal, 2000, describe formative assessment as an attitude of teaching, not only a role of feedback. Most of all, they claim, it needs well-developed observation skills and teacher's competence to provide support for students. They suggest discussions and exchange of experience

among teachers to be the easiest way to train these skills, and encourage on-the-job-training to be focussed on that aim.

Two independent groups, both working on maths assessment, have come to the same conclusion in terms of urgently needed assessment tools: Smit & Birri, 2012, found, that ready-made maths units including rubrics for assessment encourages teachers to assess complex (and therefore often neglected) competences. Jundt & Wälti, 2011, went even one step further and developed several school books with attuned assessment material at different levels of performance (details in questions 7 and 8).

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results used formatively and/or vice versa?

One research group has been writing on this issue in Switzerland: Allal, 2010, raises the question if, although formative and summative assessments have clearly different goals, possible synergy in promoting learning can be utilised. She suggests several ways to develop continuity between formative and summative assessment:

The first is through the alignment of both types of assessment with the curriculum goals underlying teaching and learning in the classroom. If this alignment is clearly perceived by students, the impact on their own goal setting can be very strong.

The second way involves high-quality feedback about learning outcomes for students: a graph comparing outcomes on different parts of a test, a set of rubrics describing the qualities of a text, or teacher comments that accompany a grade. This is thought to be helpful for regulation of student's subsequent investment in learning.

A third point of continuity has to do with student involvement in summative assessment. This form of assessment inevitably entails a judgment formulated by a professional (teacher, examiner, or other expert) about the quality of student learning. It is possible, nevertheless, to develop some degree of active student engagement in the way summative assessment is conducted. For example, in portfolio assessment used for summative purposes (grading and certification), students can participate in the selection of the work samples to include in the portfolio and be asked to write self-reflective commentaries that accompany and put into perspective their work.

On the current practice in schools regarding this continuity between formative and summative assessment, no literature could be found.

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

Related to the uptake of more formative assessment, problems on very different levels have been reported: traditional views in terms of assessment (such as: grades as results, mistakes are bad, teacher is fully responsible for assessment) have been mentioned by Rothenbacher, 2010. He agrees that this is a general problem and not limited to STM, but stresses that it is especially challenging as maths is thought to be graded objectively unlike essays etc. Jundt, 2013, adds that the main purpose of assessment

is in some cases generating grades for the school report - and not necessarily diagnosis and support for learning.

Smit, 2009, expresses the same idea more generally: that not all teachers understand the purpose of formative assessment. His studies show that teachers – and students as well - are focussed on grades, not other feedback.

On a very different level is another reason that hinders the uptake of more formative assessment: the lack of time. In teacher interviews, teachers reported to have no time to give formative feedback during classes, and to have no time to develop tools for formative assessment with teaching mates either (Smit, 2009).

In order to promote formative assessment, Smit, 2009, considers the gradual change of culture of assessment and teaching in the schools necessary. This transformation of teacher's mentality (assessment as means of supporting student's learning) should be supported from outside, by the educational system (consistent with, e.g., Assessment Reform Group, 1999). Furthermore, Smit regards the lack of knowledge in how to differentiate between several levels of proficiency in the same class as the main flaw on the side of teachers.

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

No references covering this issue in Switzerland could be found.

According to the way we see it, grounds for more IBE are there: as there are no high-stake summative assessments, teachers have a lot of freedom in conceiving their classes. This is especially true in science, less so in maths. With more support for teachers in formative and summative assessment, IBE could gain significance. Currently, IBE is not used very frequently in the Swiss teaching practice.

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

The two questions will be answered in one section, as it is difficult to split the contents. In order to provide some structure, subtitles were inserted.

There is a consistent direction among the various developments in the field of assessment, though no formal harmonisation (Vögeli - Mantovani, 1999): formative assessment is generally gaining weight and is more and more supported by regulations.

Minimal standards in the new curricula

Curriculum 21 (original name: "*Lehrplan 21*"; valid for the German-speaking region of Switzerland; will be implemented from 2015/16) and the corresponding minimal standards in science and maths education are, as discussed in question 1, geared to competency models. Many of the set standards in science and maths can be fostered with IBE. To illustrate this with an example, within the minimal standards in science, the competence "to ask questions and to investigate" (original name: "*fragen und untersuchen*") is refined in the following aspects: "perceive living creatures, situations, processes and bring up questions, problems and hypotheses", "plan and conduct investigations, surveys, or experiments", "choose suitable instruments and materials for investigations, experiments, and constructions" and "synthesise and reflect over results and methodology" (original terminology translated from EDK, 2011b). In maths, minimal standards are comparably oriented to competency models. There are some publications, in which authors explicitly link these competences with assessment tools (see paragraph on rubrics, for example).

The "assessing maths formatively" project ("*Mathematik förderorientiert beurteilen*")

This project is based on the thesis that IBE is, in practice, only possible with a corresponding assessment methodology (e.g. Rothenbacher 2012). The researchers therefore developed, tested and documented a comprehensive evaluation of student's performance with feedback on learning journals, on the student's activities, on tests, discussions about maths and student's reflections (Wälti, 2007). Their assessment instruments are explicitly designed with the aim to be integrated in daily teaching (e.g. project homepage zahlenbu.ch). In order to allow individual support, they defined different levels of performance.

The maths book series "Zahlenbuch" (for primary school levels; Berger et al., 1996; Berger et al., 1997; Berger et al., 1998; Berger et al., 1999; Affolter et al., 2000; Affolter, 2002) and "Mathematische Beurteilungsumgebungen" (for lower secondary; Jundt & Wälti, 2011; Jundt & Wälti, 2012) is closely interlinked with these assessment tools and clearly indicates ways to assess IBE-related competences such as problem solving competence.

Tendencies indicate that students at primary school level work with more motivation and endurance after learning with these maths books, but not necessarily show better performances (Rothenbacher, 2011).

Rubrics (a maths project)

Smit & Birri, 2012 developed a rubric to assess activity-oriented competences in maths. In more detail, they focussed on competence in "argumentation and reasoning" (original terminology in the newly developed educational standards in maths, EDK 2011a: "*Argumentieren und Begründen*"). The dimensions of their rubric are "adequate and understandable approach", "correct arithmetics", "understandable and detailed description and justification", "sketches and examples" (in consistency with the educational standards in maths). Along with this rubric goes a ready-made unit of about 5 lessons with exercises and tasks from official school books. Results from the implementation are not published yet. The same authors (Birri & Smit, 2013) developed a second rubric

focussing on "investigation and exploration" (original terminology in the educational standards in maths, EDK 2011a: "*Erforschen und Explorieren*") and corresponding dimensions. Again, a maths unit was planned along with the rubric. This time, the authors focussed on self-assessment by students and on the fact that students should be aware of the educational aims. Therefore, the dimensions of the rubric were formulated in easily understandable questions instead of abstract keywords.

Smit, 2009, also designed rubrics for formative assessment in self- and methodological competence and found, similar to the results in the "assess maths formatively" project, that students' self-confidence increases, though not necessarily their performance assessed by the teacher.

"Lernwelten Natur - Mensch - Mitwelt" (a science project)

A similar approach was chosen by the authors of a series of science schoolbooks (original title of the individual volumes for primary school level: "Karussell", "Riesenrad", "phänomenal"). There are strong interrelations between the books but also with the volume providing the didactical background ("Lernwelten Natur - Mensch - Mitwelt"; Adamina & Müller, 2008). All the schoolbooks promote IBE in various topics. The appending teacher comments explicitly point to situations for teacher's feedback. The volume called "Riesenrad" (Wyssen, Bringold & Kiener, 2005) for 3rd and 4th grade students shall be described in more detail to give an idea of the approach. Topics covered in this volume are day and night, magnetism, forces, air, soil, water, pond, forest, village and town and seasons. Situations to assess students' protocols (texts and sketches on presumptions, observations and explications on phenomena or experiments) as well as the oral presentation of findings can be found for all the topics. The authors indicate three more possible assessment settings: compare student's pre- and post-concepts; let them develop information autonomously and assess their result; present a new situation to students and assess their transfer from a known field.

Portfolios (no specific subject)

Portfolios have been used in various assessment situations, most often in foreign language learning (Winter, 2009). They have also been suggested to display students' efforts, progress and achievements in maths (Vögeli - Mantovani, 2002). Winter, 2009, agrees on this and promotes wider use of portfolios. He explicitly mentions their advantage compared to traditional paper-and-pencil-exams: for portfolios, relevant documents have to be collected over a longer period of time. This means that, necessarily, students are actively involved, they are much more aware of the criteria for success and assessment is therefore integrated in daily learning.

Closing remark

Before closing this section, it should be stressed that the described methods and instruments are more progressive than the average teaching practice. They do not reflect daily business at Swiss schools.

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

The schools for teacher education have not been generally doing a lot of research on formative assessment, neither have the universities.

TIMMS and PISA results have provoked discussion in the concerning circles, but assessment practice has not changed.

Question 10: Which are the 10 most relevant publications on formative and summative assessment of inquiry in science, technology, and mathematics in your country (please make sure that your selection covers all three subject areas)? If there are no publications especially dealing with IBE, please list the 10 most relevant publications in formative and summative assessment in general. What are the main results of these publications?

- Adamina, M. & Müller, H. (2008): *Lernwelten Natur-Mensch-Mitwelt. Grundlagenband zur Reihe Lern-Lehrmaterialien zum Fach NMM*. Bern: schulverlag plus.
Reference work on didactics of science and related subjects (adapted to the situation in Switzerland).
- Allal, L. (2010): Assessment and the Regulation of Learning. In P. Peterson, E. Baker, B. McGaw (Eds), *International Encyclopedia of Education (volume 3)*, 248 - 352. Oxford: Elsevier.
An article summarizing models of regulation of learning, contributions of assessment to the regulation of learning, and continuity between formative and summative assessment.
- Allal, L. & Mottier Lopez, L. (2005): *Formative Assessment of Learning: A Review of Publications in French*. Retrieved from <https://www1.oecd.org/edu/cei/35337948.pdf>
A review based on articles appearing in the journal "Mesure et évaluation en éducation".
- Birri, T. & Smit, R. (2013): Lernen mit Rubrics. Kompetenzen aufbauen und beurteilen. *Pädagogik* 3/2013, 36 - 39.
Experience from introducing and operating with rubrics (competence grid)
- EDK Swiss Conference of Cantonal Ministers of Education (2011): *Grundkompetenzen für die Mathematik. Nationale Bildungsstandards*. Retrieved from <http://www.edk.ch/dyn/11673.php>.
Swiss educational standards in maths, developed in the context of the harmonisation of compulsory school.
- EDK Swiss Conference of Cantonal Ministers of Education (2011): *Grundkompetenzen für die Naturwissenschaften. Nationale Bildungsstandards*. Retrieved from <http://www.edk.ch/dyn/11673.php>.
Swiss educational standards in science, developed in the context of the harmonisation of compulsory school.
- Jundt, W. & Wälti, B. (2011): *Mathematische Beurteilungsumgebungen*. Schulverlag Plus: Bern.
A school book; collection of tasks which assess typical competences related to maths (in dependence on the educational standards in maths)

- Rothenbacher, M. (2010): *Beurteilen im Mathematikunterricht mit dem Zahlenbuch: Gegriffs -Klärungen, -Verständnis, -Grundlagen*. Retrieved from http://www.zahlenbu.ch/cms/media/archive3/kursunterlagen_zahlenbuch/WB_Kurs_Beurteilung_MATH_Grundlagen_2010.pdf
A guide clarifying terms related to formative assessment.
- Smit, R. & Birri, T. (2012): *Lernen mit Rubrics als Teil der formativen, standard-orientierten Beurteilung*. Unpublished manuscript, PH St. Gallen.
Experience from introducing and operating with rubrics (competence grids) in maths.
- Vögeli-Mantovani, U. (1999): *SKBF Trendbericht Nr. 3: Mehr fördern, weniger auslesen. Zur Entwicklung der schulischen Beurteilung in der Schweiz*. Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.
Trend report on assessment practices in Switzerland.
- Wälti, B. (2007): *Mathematik förderorientiert beurteilen. Grundschulunterricht 8/2007, 24 - 27*.
Article promoting progress-oriented assessment of students' performance; final annual grades based on students' products (exercises), talks over these products, hands-on exercises, traditional tests, self-assessment of students.

References

- Adamina, M. (2010): *Lernen begleiten, begutachten und beurteilen*. In P. Labudde (Ed.), *Fachdidaktik Naturwissenschaft 1. -9. Schuljahr (181-196)*. Bern: Haupt Verlag.
- Affolter, W., Amstad H., Doebeli, M. (2000): *Schweizer Zahlenbuch 5*. Klett & Balmer: Zug.
- Affolter, W. (2002): *Schweizer Zahlenbuch 6*. Klett & Balmer: Zug.
- Allal, L. & Mottier Lopez, L. (2005): *Formative Assessment of Learning: A Review of Publications in French*. Retrieved from <https://www1.oecd.org/edu/cei/35337948.pdf>
- Allal, L. (2010): *Assessment and the Regulation of Learning*. In P. Peterson, E. Baker, B. McGaw (Eds), *International Encyclopedia of Education (volume 3)*, 248-352. Oxford: Elsevier.
- Assessment Reform Group (1999): *Assessment for Learning. Beyond the Black Box*. Retrieved from <http://gtcni.openrepository.com/gtcni/bitstream/2428/4621/1/Assessment%20for%20Learning-Beyond%20the%20Black%20Box.pdf>
- Berger A., Fischer M., Birnstengel-Höft, U., Wieland G., Hengartner, E. (1996): *Schweizer Zahlenbuch 1*. Klett & Balmer: Zug.
- Berger A., Fischer M., Birnstengel-Höft, U., Wieland G., Hengartner, E. (1997): *Schweizer Zahlenbuch 2*. Klett & Balmer: Zug.
- Berger A., Fischer M., Birnstengel-Höft, U., Wieland G., Hengartner, E. (1998): *Schweizer Zahlenbuch 3*. Klett & Balmer: Zug.
- Berger A., Fischer M., Birnstengel-Höft, U., Wieland G., Hengartner, E. (1999): *Schweizer Zahlenbuch 4*. Klett & Balmer: Zug.
- Birri, T. & Smit, R. (2013): *Lernen mit Rubrics. Kompetenzen aufbauen und beurteilen. Pädagogik 3/2013, 36-39*.

- Dzelili, A. (2009): Die grosse Frage im Hintergrund: Wozu ist die Schule da? In D. Fischer, A. Strittmatter, U. Vögeli-Mantovani (Eds.), *Noten, was denn sonst? Leistungsbeurteilung und -bewertung* (41-46). Verlag LCH: Zürich.
- EDK Swiss Conference of Cantonal Ministers of Education (2011a): *Grundkompetenzen für die Mathematik. Nationale Bildungsstandards*. Retrieved from <http://www.edk.ch/dyn/11673.php>.
- EDK Swiss Conference of Cantonal Ministers of Education (2011b): *Grundkompetenzen für die Naturwissenschaften. Nationale Bildungsstandards*. Retrieved from <http://www.edk.ch/dyn/11673.php>.
- Fischer, D. (2009): Keine Noten - keine Beurteilung? In D. Fischer, A. Strittmatter, U. Vögeli-Mantovani (Eds.), *Noten, was denn sonst? Leistungsbeurteilung und -bewertung* (25-26). Verlag LCH: Zürich.
- Frey, K. & Frey-Eiling, A. (2004): *Allgemeine Didaktik*. Verlag der Fachvereine Zürich: Zürich.
- Husfeld, V. (2009): Aus der Praxis der Leistungsbeurteilung. In D. Fischer, A. Strittmatter, U. Vögeli-Mantovani (Eds.), *Noten, was denn sonst? Leistungsbeurteilung und -bewertung* (33-40). Verlag LCH: Zürich.
- Jundt, W. & Wälti, B. (2011): *Mathematische Beurteilungsumgebungen Sek I/1*. Schulverlag Plus: Bern.
- Jundt, W. & Wälti, B. (2012): *Mathematische Beurteilungsumgebungen Sek I/2*. Schulverlag Plus: Bern.
- Jundt, W. (2013): Unpassendes zur Beurteilung. *profil. Magazin für das Lehren und Lernen* 1/ 2013, 10-11.
- Kronig, W. (2009): Schulnoten - Glasperlen des Bildungssystems. In D. Fischer, A. Strittmatter, U. Vögeli-Mantovani (Eds.), *Noten, was denn sonst? Leistungsbeurteilung und -bewertung* (27-32). Verlag LCH: Zürich.
- Rothenbacher, M. (2010): *Beurteilen im Mathematikunterricht mit dem Zahlenbuch: Begriffs - Klärungen, - Verständnis, - Grundlagen*. Retrieved from http://www.zahlenbu.ch/cms/media/archive3/kursunterlagen_zahlenbuch/WB_Kurs_Beurteilung_MATH_Grundlagen_2010.pdf
- Rothenbacher, M. (2011): *Projekt "Mathematik förderorientiert und ganzheitlich beurteilen. Zwischenbericht 2011*. Retrieved from http://www.zahlenbu.ch/cms/media/archive3/Projekt_MATH_beurteilen_Zwischenbericht_2011_Internetver.pdf
- Rothenbacher, M. (2012): *Förderorientiert beobachten und beurteilen im Mathematikunterricht*. Retrieved from http://www.zahlenbu.ch/cms/media/archive3/Artikel_profiL_121212_Entwurf_Langversion.pdf
- Ruf, U. & Gallin, P. (1991): *Lernen auf eigenen wegen - mit Kernideen und Reisetagebüchern*. Retrieved from http://www.bzl-online.ch/archivdownload/artikel/BZL_1991_2_248-258.pdf
- Schwartz, G. & Allal, L. (2000): *Vers une pratique de l'évaluation formative dans le secondaire I. Analyses d'expériences menées au Cycle d'orientation de Genève*. DIPCO: Genève.

- Smit, R. (2009): *Die formative Beurteilung und ihr Nutzen für die Entwicklung von Lernkompetenz. Eine empirische Studie in der Sekundarstufe 1.* Schneider Verlag Hohengehren GmbH: Baltmannsweiler.
- Smit, R. & Birri, T. (2012): *Lernen mit Rubrics als Teil der formativen, standardorientierten Beurteilung.* Unpublished manuscript, PH St. Gallen.
- Wälti, B. (2007): Mathematik förderorientiert beurteilen. *Grundschulunterricht* 8/2007, 24 - 27.
- Winter, F. (2009): Das Portfolio: Impulse für das Lernen und die Leistungsbeurteilung. In D. Fischer, A. Strittmatter, U. Vögeli-Mantovani (Eds.), *Noten, was denn sonst? Leistungsbeurteilung und -bewertung* (53-60). Verlag LCH: Zürich.
- Wyssen, H.-P., Bringold, B. & Kiener, J. (2005): *Riesenrad. Natur und Technik, 3./4. Schuljahr.* schulverlag blmv AG: Bern.
- Vögeli - Mantovani, U. (2002): *Portfolios im Mathematikunterricht der Primarschule.* Retrieved from <http://www.wolfsweb.ch/nwedk/nwedk2/index.html>
- Vögeli - Mantovani, U. (1999): *SKBF Trendbericht Nr. 3: Mehr fördern, weniger auslesen. Zur Entwicklung der schulischen Beurteilung in der Schweiz.* Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.

2.8 Report of Assessment of STM in England

Christine Harrison¹⁴

2.8.1 Introduction

This report has been constructed mainly from a search of Taylor and Francis Journals using keywords of formative assessment, assessment for learning, teacher assessment, summative assessment, science assessment, mathematics assessment and technology assessment and UK SATs. 42 papers were noted as relevant for the study questions by reading the abstracts and findings. These were reduced to 24 by selecting those referring to a school context rather than a higher education context. Other documents were consulted (eg ASL14-19 report, OfSTED Assessment for Learning Implementation report and OfSTED Assessing Pupil Progress report)

2.8.2 Answers to the questions

Question 1: Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?

Any consideration of the science curriculum and its assessment must begin with an understanding of the goals and purposes of science, technology and mathematics education. Only when these are clearly defined and recognised is it then possible to produce an assessment system that supports and augments the curriculum. In 1989, the government introduced a National Curriculum for the 5-16 age range within state schools but failed to provide aims for that curriculum.

Traditionally, the science curriculum in England has been a pre-professional preparation for the next generation of potential scientists. Such an education is, whatever people may say, more akin to a form of pre-professional training for the future scientist rather than an education about science and the insights offered by the world-view it presents. Miller & Osborne (2000) challenged this approach to teaching and learning science and called for a curriculum that addressed the needs of the scientific citizen for all students and specialist science courses for those youngsters eager for a scientific career. While there have been some moves towards this approach, much of science teaching and learning and certainly of assessment in science has remained focused on science for the future scientist.

Consequently, the curriculum begins with the foundations of science addressing basic concepts in a piecemeal fashion, which often seem unrelated to the neophyte student. On these foundations are built additional layers of knowledge. Any sense of coherence, based on an understanding of the major themes and interrelationships within the sciences, is only obtained after many years of study. In addition, the knowledge that forms the substance of such a curriculum is well established, unequivocal, seen as not available for questioning – and therefore not questioned. Consequently, there is a tendency for much of the subject to be taught in an authoritarian manner leaving little

¹⁴ Corresponding author (e-mail: christine.harrison@kcl.ac.uk)

space for discussion or exploration of what, after all, are a set of unnatural and difficult ideas.

Mathematics teaching has also followed a rather traditional approach in England with the emphasis more on calculation and algorithmic solutions than on problem solving. On the other hand, technology has embraced a design rather than the craft approach popular in the 1970s.

Summative assessment has dominated the assessment field following the introduction of the National Curriculum in 1989. The Assessment of Science Learning report commissioned by Wellcome attempted to summarise the state of assessment in school science in the UK and to suggest ways forward. The government, at the time, rejected any changes in the 14-19 curriculum and assessment framework despite the call for change in this and other papers at the time.

Question 2: How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?

The concept of formative assessment had featured in the literature for over 20 years before it was highlighted in the TGAT report (1988) which stated that it should be central to any national assessment system. In 1998, the seminal paper by Black & Wiliam, along with a booklet produced for teachers and schools, catapulted formative assessment into the assessment arena. Their review reported on the positive effects on learning where formative assessment had been used in classrooms and the paper called for innovations in practice to establish such practice in UK schools. Subsequent projects established that such practices could be established in schools (Black et al, 2002, 2003; Wiliam et al 2004) while documenting the radical changes that teachers needed to bring about to establish such practice. (Harrison, 2006, 2013).

In formative assessment, the goal is to find out what students know, what they partly know and what they do not know (Black and Harrison 2004). This awareness comes out of activities that encourage students to talk about their learning, and to apply whatever knowledge they have, from which teachers can gauge their level of understanding. The idea is to try to elicit knowledge of student understanding, and so teachers need to explore the ways in which their students are making sense of their learning experiences. At the same time, students will be able to compare their developing understandings and ideas with those of their peers (Stiggins, 2007).

The ultimate aim of formative assessment is to help students develop a self-regulated approach to learning (Harrison & Howard 2009) where students use assessment to look honestly at where they are at and utilize the assessment process to help them move forward in their learning. Formative assessment provides teachers and learners with data on learning so that future learning experiences can be matched to the learner's needs. This process supports teachers in matching the pace of learning and amount of challenge to their students. For students, formative assessment helps them develop and extend their self-assessment skills and learning behaviours to encourage them to focus on their learning and seek help through collaboration and discussion with

their peers. For this to happen, teachers need to work on student learning behaviours and in particular dialogue, collaboration and self-regulation (Harrison & Howard 2009).

Some teachers use pre-emptive formative assessment (Carless 2007) in that they plan activities or ask questions that, from experience, they are aware may cause problems in student learning. So they may do some small activities to sort out areas they feel may be problematic in a later piece of study or larger learning activity. While such an approach could be seen as scaffolding and supportive of learning, sometimes this approach provides a shortcut which enables learners to complete a difficult task but without providing opportunity to fully sort out the initial problem in the learning. It prevents students making mistakes but sometimes it is easier to learn from making a mistake and then working out how to do a task without that mistake than having strategies which avoid the mistake in the first place. We have noticed this approach especially in our SAILS project, and teachers are finding it difficult to make decisions about whether a student can perform a particular inquiry skill or not because the students have been helped overcome the difficulty within the task and teachers are uncertain whether the learners could be successful without teacher scaffolding.

In 2010, a pilot was launched nationally called Assessing Pupil Progress. Its claim was that it provided a new approach to Assessment for Learning. In fact, it was a criterion-referenced approach to assessment organized into several topics which was useful in helping teachers make judgments about National Curriculum levels. One of these topics was Science at Work, which covered some aspects of inquiry. In essence, it was a help to making summative decisions but some teachers manage to use this system in a more formative way, at least in terms of informing their planning. While reference to this system is found in inspector reports, there has been nothing published on this to date.

Question 3: What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?

Sach (2011) conducted a study in primary and middle schools with 24 teachers to ascertain their perceptions and practice in relation to formative assessment. She found a much more complex picture than she anticipated in that the nature of teachers' perceptions was strongly influenced by how these were formed, the particular school contexts and how they may impact on practice. We found similar in the KREST project (2006) and what was important was a means of developing a professional language to discuss changes that teachers implemented in their classrooms and the impact this had on learners and on themselves. In this project, we approached this through a collaborative action research approach and the construction of a reflective portfolio that enabled the teachers opportunity to explain and contextualize the change they reported. (Harrison 2013).

As mentioned earlier, teachers need help in setting up the classroom environment in which assessment for learning can flourish. Marshall and Drummond (2006) introduced the idea of 'letter or spirit of AfL' after observing teachers introducing AfL in secondary classroom contexts where they noticed some teachers often used finite activities,

closed instructional dialogue and teacher dependent exchanges which made AfL practices more like 'procedures'. The teacher's hierarchical relationship with the class ("my pupils") and the attribution of ineffectiveness placed on the students' lack of readiness rather than an ownership by the teacher to create the readiness led the teacher to do most of the work for the students. Learner autonomy was viewed as an added bonus rather than a stated aim. However, teachers who saw themselves as learners, who were keen to promote independence in students and who saw it primarily as their responsibility to initiate learning readiness, were most effective in creating the learner autonomy that they sought. In other words, 'the spirit' of AfL infiltrated all aspects of the learning scenario and was embedded in the practice. In our experience, many mathematics and science teachers require support in creating this type of classroom environment.

Question 4: How can summative assessment be used to support formative assessment? Are summative assessment methods and results used formatively and/or vice versa?

Harlen (2009) suggests that teachers need to be clear about the ways in which information gathered as part of teaching can be used formatively to help learning and then summarised and judged against reporting criteria for summative uses. This she claims requires attention from both pre-service and in-service professional development in extending teachers' understanding of learning goals, assessment criteria and moderation procedures.

Both formative and summative assessment are essential components of teaching and learning. In general terms, assessment is simply the production and interpretation of evidence of achievement. If this evidence is used to guide the next steps in progress, it is for learning; if it is used to sum up, judge, make decisions about, progress so far, it is of learning.

The King's group presented a unified view of the role of assessment in pedagogy in terms of the five stages that needs to be involved in any teaching activity, as follows (Black et al, 2013):

A Formulating Aims. This is the stage of strategic decision. All that follows should relate to a clear formulation of the learning aims.

B Planning Activities. The aims are to be achieved by choosing, adapting, or inventing activities which will engage pupils, and thereby elicit responses from them which help to clarify and then extend their understanding

C Implementation. The way in which a plan is implemented in the classroom is crucial. What is needed is formative interaction which stimulates and builds on the pupils' contributions. This is the core activity of assessment for learning.

D Review. At the end of any learning episode, there should be a review, to check before moving on. The assessment used at this stage may be designed to be summa-

tive, but its results can also be for learning, e.g. to help all pupils, through peer marking, to develop understanding of the criteria of quality in meeting the aims. It may also identify for the teacher a need to revisit some issues with the class as a whole.

E Summing Up. This is a more formal version of the *Review* stage: here the results may be used to make decisions about a pupil's future work or career, to report progress to other teachers, school managements and parents, and to report the overall achievements more widely to satisfy the need for accountability.

The overall point of this five-stage approach is that teaching, learning, and assessment should all be closely inter-linked in the planning and implementation of any teaching programme. To treat assessment as a marginal or secondary issue is to impoverish the work, by missing opportunities for improvement and by creating future tensions

Question 5: What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?

In 2004, Assessment for Learning (AfL), had been adopted as a National Strategy for whole school improvement and a range of professional development materials were produced centrally and rolled out to schools. While this move did undoubtedly raise the profile of AfL in schools nationally, implementation was somewhat sporadic in most schools. This was reported in the Assessment for Learning 8 Schools Project, which was an action research project, and resulted in 13 recommendations to improve implementation. Four of these were associated with teaching and learning while the remaining nine focused on how senior leaders could prevent competing priorities, the demands for summative data and other issues of accountability from slowing the implementation of formative practice. To make such changes is a massive undertaking by a school. Boyle & Charles (2010) also found that the summative agenda straitjacketed teaching and learning in the 43 schools they surveyed and this hampered the development of formative practice.

At classroom level, one of the main drawbacks for teachers was in fully developing the dialogic classroom. While questioning had been a major area for teachers to work on in developing their practice, still too much of the talk resided with the teacher and therefore both teacher and learners are unable to get sufficient feedback as learning is taking place to adjust their ideas in relation to the evidence.

Nevertheless, some important implications for pedagogical practice have already emerged at theoretical and practical evidential levels and these support recommendations for future practice. This includes the need for teachers to have an in-depth pedagogical knowledge of how children learn and of their own pupils' learning needs (Shulman 1986; Watkins and Mortimore 1999). This would ensure that assessment practices encourage deep, rather than superficial, knowledge and understanding. Allowing pupils to strengthen their self- and peer-assessments through collaborative endeavor so that they come to share responsibility for their own learning is another important facet. The assessment of progress against personal rather than normative frameworks is a further key issue here, supported by Dweck's theory (2000) of

achievement motivation.

Question 6: Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?

Certainly the emphasis on coursework at GCSE has greatly hampered IBE development. Much of the coursework done in schools in all three subject areas is ritualistic and narrow in focus. Teachers are with their students for extensive periods of time, constantly interacting with them inside and sometimes outside the classroom, posing questions and noting responses, and observing performances as they carry out assigned tasks and activities. However, many teachers are reluctant to do teacher assessment when it requires them to report. An important motivation for using teacher assessment in place of, or in addition to, tests is to recognise, develop and value the professionalism of teachers. One negative impact of the higher profile given to test-based results in England's national curriculum assessment system has been shown to be not only a loss of assessment skill on the part of teachers, but also a loss of confidence in their ability to make sound assessments of their students (Black et al. 2010, 2011). Another factor, supported by the teacher unions, is concern over time and workload.

Question 7: Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?

As mentioned earlier, some teachers utilize aspects of the Assessing Pupil Progress framework relatively effectively to decide on National Curriculum levels and, in some cases, to inform their planning.

Question 8: Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?

Robin Millar and his team at York are developing diagnostic questions in science currently which may be of use formatively. The King's group have produced a series of subject specific booklets to support the development of formative practice. Some commercial publishers have produced multiple choice questions which are used with the electronic whiteboard and students can vote on the option they feel is correct. Most of these do not have guidance on how to use these results formatively, but some teachers are working this out and using them in this way. Equally, some teachers use Socrativ in a similar way with mobile phones. While this is very new in classrooms, there is a keenness for this approach to develop further.

Question 9: How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

Certainly the legacy of the Black & William 1998 review can be seen throughout schools in that most teachers and schools are aware of what assessment for learning is.

The government uses snippets from the many international surveys to justify some of their decisions with regard to schools and their requirements. Because of the high stake hold on schools in that results, positions in league tables and reaching targets are seen as of paramount importance, research is acknowledged but often ignored or selectively used by schools and teachers.

Question 10: Which are the 10 most relevant publications on formative and summative assessment of inquiry in science, technology, and mathematics in your country (please make sure that your selection covers all three subject areas)? If there are no publications especially dealing with IBE, please list the 10 most relevant publications in formative and summative assessment in general. What are the main results of these publications?

Paul Black, Chris Harrison, Jonathan Osborne, Rick Duschl (2006) Assessment of Science Learning 14-19 Report. Wellcome: London.

Paul Black , Christine Harrison , Jeremy Hodgen , Bethan Marshall & Natasha Serret (2010) Validity in teachers' summative assessments, Assessment in Education: Principles, Policy & Practice, 17:2, 215-232,

Paul Black , Christine Harrison , Jeremy Hodgen , Bethan Marshall & Natasha Serret (2011) Can teachers' summative assessments produce dependable results and also enhance classroom learning?, Assessment in Education: Principles, Policy & Practice, 18:4, 451-469,

William Francis Boyle & Marie Charles (2010) Leading learning through Assessment for Learning?, School Leadership & Management: Formerly School Organisation, 30:3, 285-300,

David Carless (2007) Conceptualizing pre-emptive formative assessment, Assessment in Education: Principles, Policy & Practice, 14:2, 171-184,

Wynne Harlen (2005) Teachers' summative practices and assessment for learning – tensions and synergies, Curriculum Journal, 16:2, 207-223,

Chris Harrison & Sally Howard (2009) Inside the Primary Black Box. GLAssessment: London

Christine Harrison (2013) Collaborative action research as a tool for generating formative feedback on teachers' classroom assessment practice: the KREST project, Teachers and Teaching: Theory and Practice, 19:2, 202-213,

Marshall, Bethan and Drummond, Mary Jane (2006)'How teachers engage with Assessment for Learning: lessons from the classroom',*Research Papers in Education*,21:2,133 — 149

Dylan Wiliam, Clare Lee, Chris Harrison & Paul Black (2004) Teachers developing assessment for learning: impact on student achievement *Assessment in Education, Vol. 11, No. 1, March 2004 1-18*

Annex – Questions to guide national collection of research (D2.1)

In answering the following questions, please keep in mind that we are interested in research results and not in information at the level of the educational systems in your country! This information will be collected by WP3!

With respect to the research on assessment in your country, what do the results suggest regarding the following questions (please include citations)?

1. Which role does summative and formative assessment play in and for the teaching and learning of STM in your country?
2. How do teachers approach the need to monitor student learning as it develops? To what extent do they use structured formative assessment and in what formats?
3. What support and tools do teachers need in order to integrate formative assessment of student learning in their classroom practice?
4. How can summative assessment be used to support formative assessment? Are summative assessment methods and results formatively and/or vice versa?
5. What hinders the uptake of more formative assessment? What need to be done to promote formative assessment?
6. Do the assessment methods influence the uptake of IBE in STM in your country? How can summative and formative assessment methods be used to promote learning in inquiry-based STM?
7. Is there evidence that summative or formative assessment methods in your country exist that measure the goals of IBE in STM: understanding of powerful scientific and mathematical ideas, building inquiry competences, developing understanding of scientific and mathematical activity and fostering corresponding attitudes? What results are found? If the assessment methods are not consistent with the learning goals of IBE, how could they be brought into consistency?
8. Do instruments and protocols (including ICT) for formative assessment exist? If yes, which and how are they used?
9. How are research results on assessment used in your country and by whom? Which significance do they have at the student level, classroom level, teacher level (or maybe beyond teacher level)?

10. Which are the 10 most relevant publications on formative and summative assessment of inquiry in science, technology, and mathematics in your country (please make sure that your selection covers all three subject areas)? If there are no publications especially dealing with IBE, please list the 10 most relevant publications in formative and summative assessment in general. What are the main results of these publications?