www.sails-project.eu

Report on the strategy for the assessment of skills and competencies suitable for IBSE

SAILS

Strategies for Assessment of Inquiry Learning in Science

Strategies for Assessment of Inquiry Learning in Science

D 2.1 Report on the strategy for the assessment of skills and competencies suitable for IBSE

Authors: Benő Csapó, Csaba Csíkos, Erzsébet Korom, Mária B. Németh, Paul Black, Christine Harrison, Paul van Kampen and Odilla Finlayson
Project name: Strategies for the Assessment of Inquiry Learning in Science (SAILS)
Project number: 289085
Start date: 01/01/2012
Duration: 48 months
Lead partner for this deliverable: University of Szeged
Project coordinator: Dublin City University
Contact: info@sails-project.eu
Website: www.sails-project.eu

The research leading to these results has received funding from the European Union's Seventh Framework Programme for research technological development and demonstration under grant agreement no 289085
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Contents

1 Introduction

In this document we outline the assessment strategies applicable in the context of Inquiry Based Science Education (IBSE), with a strong emphasis on diagnostic testing. An outline of non-diagnostic testing strategies is given in D3.1. We build on the results of Work Package 1, especially on the document summarizing the goals of ISBE projects (D1.1). First, we characterize IBSE from the perspective of the *Strategies for Assessment of Inquiry Learning in Science* (SAILS) project, looking at it from the point of view of assessment. In the main part, we devise taxonomy for classifying the objectives to be assessed in the context of IBSE. We assume that a variety of aims have been pursued in the different IBSE implementations, and a unified set of them don't differ much from the goals of efficient science education in general; therefore we apply the broadest possible approach to accommodate every relevant objective of science education in this taxonomy. On the other hand, we pay special attention to the particular goals of IBSE by devising a taxonomy which helps highlight the characteristic attributes of IBSE. Next, we outline the assessment strategies applicable to SAILS, and describe contexts, methods and tools for carrying out the possible strategies. One of the main functions of this document is to bridge the gap that can be observed between the expectations outlined at policy level, and the achievable goals and aims of IBSE projects. Furthermore, these goals are conceptualized in terms of assessable outcomes of learning. The document deals with the most general conceptual and strategic assessment issues related to IBSE, specified to the particular activities of SAILS, providing a foundation for the tasks of other working groups. It also identifies the main directions for the next task to be carried out within Work Package 2, the development of frameworks of assessment. Due to this segmentation of the activities, this document identifies only the main objectives of assessment, but a detailed description of them will be given in the assessment framework.

2 Brief Summary of IBSE from the Perspectives of Assessments in SAILS

In the past decades, IBSE has become one on the most prominent alternatives to traditional science education. Its popularity has generated a great variety of implementations in terms of interpretation of inquiry, depth of changes to traditional teaching methods, areas of application, complexity of inquiries, and lengths or frequency of the application of the related activities. The communication about it ranges from policy documents to academic publications reporting experimental results. In this section, we narrow the scope of the sources of information usable in the context of SAILS and identify the aspects of IBSE where assessment is relevant and possible.

2.1 Origin, Goals and Developments in IBSE

Since the beginning of organized schooling, several aims were attributed to science education. Principal among those were (1) transmitting core knowledge scientific research created, thus giving a foundation for the studies of those few who choose a science-related career, (2) to provide every citizen with the scientific literacy necessary in modern societies depending on scientific discoveries and technological innovations, and (3) to develop students' general thinking capabilities, in particular deductive reasoning skills to complement the inductive skills developed in mathematics (DeBoer, 1991).

By the middle of 20th century, several phenomena indicated that in the developed countries science education did not meet these expectations (e.g. the related studies in the US after the 'Sputnik Shock'). Science education, at best, resulted in inert knowledge; students mastered science learning materials without understanding, but were not able to apply their knowledge in new, unfamiliar contexts. Several years of learning sciences left students' naïve ideas and misconceptions untouched, and traditional teaching had little impact on their thinking skills. International student assessments provided quantitative proofs of the deficiencies of science education from the 1970s (e.g. Comber &

Keeves, 1973), and analyses of teaching materials showed that they were overwhelmed by facts and figures that are far more abstract than what can be comprehended at the actual cognitive level of students (e.g. Shayer & Adey, 1989).
Improving science education was one of the focal aims of the reform movements form the 1980s, and IBSE became a dominant alternative from the mid 1990s (Furtak, Seidel, Iverson & Briggs, 2012). In the past decades, IBSE has become popular in several countries around the World, and it was introduced in a number of ways into primary (see e.g. Harris & Rooks, 2010) and secondary (see e.g. Llewellyn, 2005) level science education. In Europe, the policy measures after the Rocard report (Rocard et al, 2007) launched a number of large-scale national and cross-national projects (see D1.1). Implementations of IBSE incorporated several attributes of the previous progressive methods encouraging students' activities, such as project method, collaborative learning, problem-based learning etc., and the common goals and methods make them sometime indistinguishable.

The proliferation of inquiry methods has generated controversial effects, too. On the one hand, the lack of a clear definition of inquiry eroded the conception, allowing too many weak implementations. In many cases, the amount of inquiry did not reach a critical threshold of effectiveness. On the other hand, results of inquiry based programs were not always convincing enough. Inquiries often require special equipment and more time than direct instruction, distracting resources from other areas and methods, causing other problems in this way. Results of studies assessing the additional benefits of inquiry-based learning compared to traditional teaching were synthesized in several meta-analyses (Weinstein, Boulanger & Walberg, 1982; Bredderman, 1983; Shymansky, Hedges & Woodworth, 1990; Wise & Okey, 1983; Lott, 1983, Schroeder, Schott, Tolson, Huang & Lee, 2007; Alfieri, Brooks, Aldrich & Tenenbaum, 2011; Gee & Wong, 2012). These analyses reported mixed results, and only under well controlled conditions have found positive effects, although as Furtak, Seidel, Iverson and Briggs (2012) have pointed out, the empirical studies conceptualized inquiry in different ways. An often mentioned concern is that without scaffolding and teacher guidance discovery learning in itself doesn't result in well-structured knowledge (Anderson, 2002; Mayer, 2004).

Many studies that have been carried out to examine the efficiency of IBSE involve experimental or quasi-experimental design, using pre- and post-tests to measure the gain resulting from the inquiry-based learning. Some approaches to measurement applied in these projects can be utilized in the context of SAILS as well, but a major difference is that the tests assessing the efficiency of inquiry methods generally (1) focus only on one or a few aspects of inquiry methods, and (2) are usually summative tests, while in SAILS, formative assessment embedded in teaching-learning processes is more relevant.

2.2 Interpretation of IBSE in SAILS, Possibilities and Challenges for Assessment

An aphorism from Einstein aptly expresses the challenges we face when considering the assessment possibilities in IBSE: *"Not everything important is measurable and not everything measurable is important."* As the goals associated with IBSE are often expressed in the form of general statements, we have to bridge the gap between policy documents and research literature. We have to translate policy statements and goals outlined at a general level into scientifically sound conceptions and have to identify the relevant and assessable constructs.

Rapid changes in post-industrial societies require new types of knowledge from their citizens so that they can be successful in private life and at work. Around the turn of the millennium, a number of projects were launched to conceptualize the knowledge the new generations are expected to possess. Deliverable 1.1 has analyzed documents dealing with these so-called 21st century skills.

Some of the most frequently mentioned conceptions in these documents are: creativity, innovation skills, critical thinking, problem solving, communication in the mother tongue and in foreign languages, collaboration and teamwork, social and cross-cultural skills, civic competencies, ICT literacy (digital competency), flexibility, adaptability, initiative and self-direction, productivity, accountability, leadership, responsibility, mathematical competency, competencies in science and technology, personal organization and time management, and learning to learn. From an assessment point of view, these conceptions are rather diverse. If diagnostic assessment is to be applied to the problem, first we have to explore if these conceptions are measurable. Measurability in this context means that there exist scientifically established instruments, or at least that the constructs to be measured are well defined – as devising instruments for entirely new constructs is beyond the scope of the SAILS project. For example, problem solving and ICT literacy are well researched domains, several assessment instruments are available, and new instruments can be constructed on the basis of the related research results. In contrast, productivity is an important personal trait, but would be difficult to assess in general, or in the context of IBSE. On the other hand, communication skills in the mother tongue and in foreign languages are well measurable, and it is conceivable that inquiries contribute to the development of these skills; their assessment in the context of IBSE is not primarily relevant.

The available instruments have mostly been prepared for summative assessments and the related research has been carried out in contexts different from supporting students' learning by frequent feedback. Furthermore, we have to take into account that the competencies they measure are the result of a long period of development. Therefore, the effects of a short period of learning on their development may not be easily detected. The consequence of these conditions is that formative diagnostic assessment in the context of IBSE is limited to dealing with skills connectable with learning taking place through inquiry.

In the context of science education, inquiry has several interpretations. Deliverable 1.1 identifies three distinct activities called inquiry: (1) what scientists do, (2) how students learn, and (3) a pedagogical approach that teachers employ. As in that document we focus on the outcomes of learning, we are interested in the second activity, and we have to identify knowledge and skills students gain when they are engaged in inquires and/or teachers employ pedagogical approaches called inquiry methods. From the perspective of SAILS, outcomes that are different or supplementary compared to other ways of learning science are especially relevant .

To make the specific outcomes of IBSE assessable, we need a clear definition of the objectives, and then we have to elaborate and describe them in detail in an operationalized format. Next we have to conceptualize the position of assessment in IBSE identifying those contexts and situations where assessment may support the learning of science and meet the given objectives. We also have to determine assessment purposes and uses in the context of SAILS. In line with the goals of SAILS, the main purpose of the assessment is to promote learning; therefore we have to integrate assessment with teaching and learning.

3 Taxonomy of Objectives to Assess in the Contexts of IBSE

As the previous sections indicated, IBSE was launched on the assumptions that if students work actively while learning science and follow the general processes of scientific discovery, then they will acquire more and better knowledge and skills. It is also assumed that continuous inquiry activities maintain interest and motivation. These are plausible assumptions, as a number of psychological and pedagogical principles suggest that active learning through inquiry may be more challenging than passively mastering the learning materials. On the other hand, less effort has been devoted to identifying and describing the particular goals of IBSE in measurable form.

Bloom and his co-workers developed taxonomic systems to describe the goals of teaching objectively. They distinguished three main domains: cognitive, affective and psychomotor. The taxonomy devised for describing the objectives for cognitive domain became the most influential (Bloom et al., 1956). These early taxonomies described the objectives of teaching in observable, behavior categories. Later, the conception of taxonomy was further developed taking into account the advances of cognitive sciences (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007).
The content of teaching and assessments may also be described by standards and frameworks. Standards focus on teaching and learning by setting the goals to be achieved at certain grades (see e.g., Waddington, Nentwig & Schanze, 2007; Ainsworth, 2003; Marzano & Haystead, 2008; O'Neill & Stansbury, 2000; Ainsworth & Viegut, 2006). Frameworks usually specify the content of assessment (e.g. OECD, 2003a, 2006, 2009a, 2013; Csapó & Szabó, 2012).

From the tradition established by Bloom's works, we preserve the distinction between the cognitive and affective domains. Although the psychomotor domain may also be relevant in certain fields of science (e.g. fine movements in some laboratory work), and a set of psychomotor objectives of IBSE could be identified, we do not deal with this domain. In this document we enumerate the assessment content, taking into account the specific features of IBSE. Framework development will be carried out in the next phase.

Several terms are used to identify the goals of education. These terms are interpreted in different ways depending on context and fields of research (e.g. knowledge, abilities, skills, competencies etc.) and bear different connotations in different languages. *Knowledge* is the most general term associated with the cognitive outcomes, but it has two meanings. It may be a unifying concept including every particular component (e.g. knowledge-economy), or it may refer only to the declarative components, as indicated by the usage of the expression 'knowledge and skills', where knowledge is complementary to skills. In this document we follow the tradition of cognitive science by using knowledge in its broader meaning (see e.g. Reif, 2008). To identify all relevant assessment objectives we use the term learning outcomes (see e.g. Bernholt, Neumann & Nentwig, 2012).

A number of different conceptual approaches were developed in the past decades to classify and present the goals of schooling and learning outcomes in a systematic way. Among those, the taxonomies, inspired by the original works of Bloom and some of their modern versions have been the most influential.

3.1 Cognitive Outcomes

In the context of SAILS, the formative approach is dominant, and assessment is embedded in the teaching-learning processes. Therefore, a detailed framework is required to elaborate the targeted learning outcomes embedded in the content of learning.

In modern societies, the same three categories of goals of science education remain relevant., However, the priorities have changed. As discussed earlier, a principal goal of science education that is supposed to be achieved especially well by IBSE approaches is (1) to develop students' general cognitive capabilities. A second goal, which has been highlighted by the PISA studies as well, is (2) to provide every citizen with scientific literacy, knowledge and skills applicable in several contexts of the developed societies. Finally, (3) transmitting disciplinary content knowledge, organized according to the principles and logic of the particular scientific disciplines is important; not only for those few who will pursue a career in science research, but for all students so they may understand the special status of proven and organized scientific knowledge. Different stakeholders place different emphasis on these goals, but inquiry-based science education can balance the three sets of goals and can contribute to attaining each.

Taking these three sets of goals into account, we propose a system in which the cognitive goals of learning science are organized in three dimensions. The theoretical background of this approach has been described in several papers (e.g. Csapó, 2012), and has already been applied in developing a diagnostic assessment framework for primary education both in science (Csapó & Szabó, 2002) and in mathematics (Csapó & Szendrei, 2002). In the next sections we describe these three different sets of goals.

3.1.1 Improving Reasoning Skills

One of the main assessable objectives of IBSE is improving students' thinking abilities, as inquiries require more intensive reasoning than traditional science education. There are several research traditions and educational approaches that focus on improving students' thinking skills and fostering cognitive development. Some successful approaches utilize regular teaching materials and modified learning activities to enhance cognitive abilities (see Csapó, 1999). The most well-known and best documented among these approaches is the Cognitive Acceleration through Science Education (CASE, see Adey, 1992, 1998; Adey, Nagy, Robertson, Serret & Wadsworth, 2003; Adey & Shayer, 1993, 1994; Adey, Shayer & Yates, 2001; Shayer, 1999). Theoretical foundations of CASE are especially relevant for establishing the psychological grounding of IBSE.

In the context of inquiry learning, reasoning skills to be developed may be organized into three major groups (see also Johnson-Laird, 2006), taking into account their complexity and their relationship to science education. These groups are (1) basic reasoning skills, (2) higher order thinking skills, and (3) scientific reasoning. These groups are not distinct, as simpler skills are embedded in more complex thinking processes.

3.1.1.1 Basic Reasoning Skills, Component Skills

In the first group, those basic reasoning skills will be discussed which have a clear operational structure, often mathematically described; some of them are even subjects of mathematics teaching (logical operations, relations, probability and randomness, combinatorial enumerations, set operations etc.). These reasoning skills are easy to describe, operationalize and assess diagnostically. The development of such basic reasoning skills was first studied by Piaget, and he often used scientific phenomena in his experiments (Inhelder & Piaget, 1958). In his studies, children were required to talk about simple mechanical experimental settings (e.g. a pendulum), identify the relevant variables and explore the relationships between them. This type of reasoning is often called *Piagetian reasoning* (Caroll, 1993) or *operational reasoning* (Csapó, 1992).

If the aim is the development of formative diagnostic assessment instruments, it has to be explored how reasoning develops and what the precursors of more complex skills are. Formative diagnostic assessment requires an analytical approach; therefore the reasoning skills relevant to mastering, organizing and applying scientific knowledge have to be identified. There are a limited number of schemas, structures, and operations which are essential in several areas of school learning and everyday problem solving. Development of these skills is essential as they play a central role in the structural transfer of knowledge between different domains.

These reasoning skills may be easily recognized in science inquiry activities. Many inquiry tasks start with the identification of variables, and continue with analyzing the relationships between them (Kuhn, Pease & Wirkala, 2009). Controlling and manipulating variables, and examining their dependencies often requires combinatorial reasoning (Kishta, 1979; Schröder, Bödeker, Edelstein & Teo, 2000; Lockwood, 2013), while organization of observation and experimental data takes place by exercising seriation, class inclusion, classification, multiple classification, set operations etc. skills. Operations of binary logic ('and', 'or', 'if … then' etc.) are essential in composing complex statements, reading and interpreting science texts, and inductive and deductive reasoning. Students' logical reasoning ability is usually a good predictor of their success in learning sciences (Leighton, 2006; Bird, 2010).

Another group of basic reasoning skills essential in science inquiries is related to the concept of probability. Some scientific phenomena are understood in terms of chance and randomness. Probabilistic reasoning (Jones, Langrall, Thornton & Mogill, 1997, 1999; Girotto & Gonzalez, 2008) is essential in risk estimation and is a precondition of the development of correlational reasoning (Lawson, Adi & Karplus, 1979; Kuhn, Phelps & Walters, 1985; Ross & Cousins, 1993; Schröder, Bödeker, Edelstein & Teo, 2000) and statistical thinking (Chance, 2002).

From the beginning of learning science, students deal with relations; thus relational reasoning is essential in understanding basic spatial and temporal relations. Linear relationships represent the most frequent connections between variables, and students have to understand the concept of ratio (Jitendra, Star, Starosta, Leh, Sood, Caskie, Hughes & Mack, 2009). Proportional reasoning (Kishta, 1979; Boyera, Levinea & Huttenlochera, 2008; Schröder, Bödeker, Edelstein & Teo, 2000) can be developed in several areas of science where linear relationships are involved (Misailidou & Williams, 2003; Taylor & Gail, 2009).

3.1.1.2 Higher Order Thinking Skills

This category comprises complex thinking processes that are often composed of simpler thinking skills (see Williams, 1999). The structure of higher order skills cannot be easily described, and different instruments may be used for their assessment. Although only a limited number of such thinking skills can be identified, they are essential in learning sciences, discovering relationships and creating new knowledge.

Analogical reasoning is the most well-known thinking process, often applied in learning science, when learners apply their knowledge mastered in one context to a new, different, but somehow (e.g. structurally) similar situation. In analogical reasoning processes, students establish a correspondence between the source (which consists of situations, structures, relationships etc. already understood), and the target, which consists of the new phenomena to be understood. Thus analogical reasoning is not only a means of learning, but is also essential in applying (scientific) knowledge to new contexts (Polya, 1968; Klauer, 1989; Abdellatif, Cummings & Maddux, 2008).

Inductive reasoning is very similar to analogical reasoning, so that some approaches consider analogical reasoning as part of the broader inductive processes (see Polya, 1968; Hamers, de Koning & Sijtsma, 1998; Csapó, 1997). Many intelligence tests are based on inductive reasoning tasks (e.g., matrix completion, continuing series), indicating the central role induction plays in human cognitive processes. Other approaches describe the basic processes of inductive reasoning as identifying similarities and dissimilarities of objects and their attributes; this letter approaches are useful for designing exercises for enhancing inductive reasoning (Klauer, 1989; Klauer & Phye, 1994, 2008), especially in young age (Molnár, 2011). Teaching methods that efficiently contribute to the improvement of inductive reasoning have been shown to develop general cognitive abilities (Shayer & Adey, 2002; McGuinness, 2005; Adey, Csapó, Demteriou, Hautamäki & Shayer, 2007).

Problem solving is a large field of research in itself. Even the term ‘problem solving’ has several meanings. In English, it is used in an especially broad sense, as ‘problems’ includes tasks, which can be solved by following simple algorithms (e.g. basic mathematical world problems) as well as difficult ones with no known solution. In other languages ‘problems’ represent complex tasks which appear in non-transparent situations, where the solution cannot be immediately seen. These complex problems require merging information from different sources, and often additional exploration as well (Frensch & Funke, 1985). Domain-specific problem solving is involved in most assessments in mathematics and science, while complex problem solving requires specific tasks. Although there is a close relationship between domain specific and complex problem solving, complex problem solving requires additional skills to comprehend novel, often dynamically changing situations, and represents ill-structured, non-transparent problems (Wirth & Klieme, 2003; Funke, 2010; Fischer, Greiff & Funke, 2012; Wüstenberg, Greiff & Funke, 2012; Schweizer, Wüstenberg & Greiff, 2013; Molnár, Greiff & Csapó, 2013).

Recently, problem solving has become a significant part of large-scale assessment projects, both at national level such as NAEP (Bennett, Persky, Weiss & Jenkins, 2007) and internationally. It was measured twice in PISA: in 2003 students solved a paper and pencil static test (OECD, 2003a, 2004), while a computer-based dynamic problem solving test represented the innovative assessment domain in 2012 (Greiff, Wüstenberg & Funke, 2012; OECD, 2013). The next PISA assessment in 2015 aims at measuring collaborative problem solving (OECD, 2012). Problem solving is the focus of the other OECD assessment program as well: the Program for International Assessment of Adult Competencies (PIAAC) assesses problem solving skills in a technology-rich environment, and the tests are administered via personal computers (OECD, 2009b).

There are a lot of similarities between scientific inquiries and problem solving. In his classical work, Polya proposed four steps to be followed in a problem solving process: (1) understanding the problem, (2) devising a plan, (3) carrying out the plan, and (4) looking back (checking, reviewing) (Polya, 1945). These steps may be identified in inquiry learning processes as well, and depending on the nature of the inquiry tasks, several problem solving skills can be applied and practiced.

Critical thinking, although often related to scientific investigation, and broadly studied, lacks a clear and generally accepted definition. When the term is used in a broader meaning, a disciplined thinking process is meant that includes systematic use of several reasoning skills described earlier, to organize an argument or to establish firm foundations for a judgment. In a narrower sense, 'critical' emphasizes reflective aspects of thinking, such as questioning and looking for proofs and evidence (Norris & Ennis, 1989; Ennis, 1995). If the definitions, descriptions, and tests of critical thinking are analyzed and the skills appearing in these descriptions are identified, no skills can be found which uniquely characterize critical thinking. Specific cognitive processes are not the distinguishing characteristics of critical thinking, but rather a critical approach, habit or attitude. Many measurement instruments (e.g. the California Critical Thinking Disposition Inventory) also deal with beliefs, values and expectations. Science projects offer possibilities to develop these critical dispositions (see Aktamış & Yenice, 2010; Eklöf, 2013).

Learning to learn is one of the European key competencies required for lifelong learning, and is defined as "the ability to pursue and persist in learning, to organize one's own learning, including through effective management of time and information, both individually and in groups" (Education Council and European Commission, 2006). Its assessment was first proposed, and its assessment framework was developed by a group of Finnish researches (Hautamäki et al, 2002). It is a comprehensive collection of skills and disposition necessary to learn effectively. The conception was extended and interpreted for international assessments (Hoskins & Fredriksson, 2008), and assessment instruments were piloted in a multinational project (Kupiainen, Hautamäki & Rantanen, 2008). Self-regulated learning (McCaslin & Hickey, 2001) is a similar construct and has also been assessed in and international project. PISA 2000 assessed students' learning habits and their approaches to learning (OECD, 2003b). As IBSE promotes independent learning, it may foster leaning to learn dispositions.

Creativity and divergent thinking have a long history of research and its assessment is dated back to the beginning of the 20th century. Creating new knowledge, discovering novel and original relationships is the essence of genuine scientific research. Learning methods based on student's active discovery processes may be the best tools for developing creativity. Learner's autonomy and peer collaboration were found important factors in supporting creative skills development (Davies, Jindal-Snape, Collier, Digby, Hay & Howe, 2013), and these aspects may be emphasized in IBSE projects as well. A number of recent studies explored the relationship between creativity and scientific activities, and identified the underlying thinking mechanisms (e.g. combinatorial reasoning, see Simonton, 2010, 2012). On the other hand, creativity is a complex construct and no significant effect may be expected from short-term interventions. Therefore, in the context of IBSE, either its component skill may be assessed, or it can be a subject of summative assessments, as well as most of the higher order thinking skills.

3.1.1.3 Scientific Reasoning

Scientific reasoning is often referred to as the most advanced form of human thinking. Scientific reasoning uses abstractions and symbols, and represents phenomena in variables and dimensions. Scientific reasoning analyses the relations between the identified symbols and variables, and in these analyses, reasoning skills described in previous sections are applied. For example, scientific reasoning often deals with ratios, proportions and probabilities; designing experiments requires systematic combination of variables involved. Argumentation requires organization of facts and figures, carrying out logical operations and establishing causal relationships between observed changes, inductive and deductive reasoning is involved (Watters & English, 1995).

Thus, the most characteristics attributes of scientific reasoning are not its building blocks, rather the rigorous and disciplined manner as they are organized, combined and executed. Most often scientific reasoning processes are long and specifically organized sequences of simpler reasoning skills.
These reasoning processes are characteristic for scientific research and learning sciences through inquiry methods alike. The basic skill which is a precursor of the other more advanced scientific thinking skills is the control and manipulation of variables. Hypothesis generation and hypothesis testing, designing experiments and analyzing results are the most typical scientific reasoning processes (Adey, 1998; Howson & Urbach, 1996; Koerber, Sodian, Thoermer & Nett, 2005; Venville, Adey, Larkin & Robertson, 2003).

Several frameworks and instruments have been developed for assessing scientific reasoning (Amsel, et al, 2008, Kind, 2013). Some of these focus at specific aspects of science learning. For example, Liu, Lee & Linn (2010) investigates how inquiries promote deeper understanding and knowledge integration. Russ, Coffey, Hammer & Hutchison (2009) assess in classroom context of how students establish causal connections in explaining natural phenomena.

One of them most elaborated conception for assessing scientific reasoning is the Evidence-Based Reasoning Assessment System (EBRAS). It focuses on how students can use evidences to support arguments (Brown, Nagashima, Fu, Timms & Wilson, 2010).

3.1.2 Improving Scientific Literacy

3.1.2.1 Improving the Transfer of Scientific Knowledge

The principles and objectives of scientific education have been described by the term scientific/science literacy. However, the interpretation of the concept varies to a great extent both in details and complexity. The literacy concepts used in practice/everyday life are unique both in terms of employing and interpreting different concepts and defining objectives. Nevertheless, the scientific literacy frameworks and standards bearing different objectives and relying on the traditions of a particular culture and education system exhibit several similar features (Aikenhead, 2007; DeBoer, 2000; Laugksch; 2000; Roberts, 2007). For example, scientific literacy is commonly considered to entail much more than the integration of knowledge, values and the fundamental elements of scientific education, as it is a complex and multi-dimensional knowledge structure (Roberts, 2007). There is a broad consensus that scientific literacy is science knowledge bearing both individual and social aspects. It is an entity applicable in everyday contexts. Most of the definitions use scientific literacy as a synonym for the slogan “public understanding of science” (Durant, 1994), implying what a scientifically literate individual/person should know about science (Roberts, 2007).

The scientific literacy assessment and curriculum frameworks bearing different objectives, perspectives, structures etc. and relying on the traditions of the local culture and education system essentially have the same building blocks (Hur, 2003, Jenkins, 1994). The components of scientific literacy described in accordance with the same aspects with different emphasis are as follows:

- content of knowledge (knowledge about relevant facts, concepts, processes and methods),
- the scientific forms of thinking/reasoning and understanding and/or the competences needed for its application,
- recognizing the values, characteristic features, objectives and limitations of sciences,
- the context of the application of knowledge (e.g. everyday/realistic, new/unfamiliar situations/contexts; or as PISA frames it: social, historical, cultural and global problems), in which the individual is expected to apply knowledge,
- interest in and attitudes towards sciences (Hur, 2003; OECD, 2006).

According to previous research and experience, in order to be able to solve the complex problems of our dynamically changing world, individuals need a transferable, expandable and adaptive science knowledge/scientific literacy. However, traditional classroom practices do not seem to be efficient in facilitating these aspects. The latest theoretical models and empirical evidences (Zhao, 2012) imply that research-based teaching methods such as Inquiry Based Science Education – IBSE may play a key role in acquiring scientific literacy, and science knowledge applicable in everyday life, and in recognizing the values and forming the attitudes towards sciences and the environment. We can also assume the positive impact Inquiry Based Science Education may have in forming scientific literacy without a detailed analysis as its core objectives meet the requirements of literacy theories and models and that of different standards.

One of the major goals of IBSE is to develop the need for acquiring knowledge, the ability for individual learning and critical reasoning (Lee et al, 2004) which contributes to acquiring Structural Scientific and Technological Literacy when students are interested in the study of a scientific concept and construct appropriate meaning of the concept from experiences (UNESCO, 2001), and extend science education throughout his or her life (Hurd, 1998). IBSE contributes to the acquisition of scientific literacy which enables individuals to engage intelligently in public discourse regarding sciences and the validity of results, critical thinking, judging the value of scientific information based on the source of the information (Lederman & Lederman, 2007; Hackling, Goodrum & Rennie 2001; MCEETYA, 2006; NRC, 1996).

If we simply consider IBSE to be a model for information processing in the teaching-learning process or the research bpased strategy for teaching and learning (Lane, 2007), one of its core elements is using/handling information. Learning the methods and strategies for the acquisition and processing (selecting) of information facilitates the ability to use scientific vocabulary, language and extensive conceptual systems in different contexts, to communicate fluently (Shamos, 1995) to evolve into functional scientific literacy (Bybee, 1997a).

The active use of the methods of scientific research and inquiry plays a direct role in the formation of scientific literacy facilitating problem solving and evidence-based decision making (Hurd, 1998). IBSE helps to acquire literacy components described in PISA, such as understanding the characteristic features, limits and impacts of science, identifying scientific issues/questions, conceptualizing evidence-based inferences (OECD, 2006). The active participation in acquiring knowledge improves the foundations of scientific literacy, a structured system of thinking processes, and the competencies needed for application by operating several abilities and skills (Hurd, 2003; Jenkins, 1994).

As students are active participants in understanding scientific phenomena, their success and failure are accompanied by emotional reactions, which are the foundations of the attitudes and environmental awareness of scientifically literate individual who experience the richness and excitement of knowing about and understanding the natural word (Lederman & Lederman, 2007; OECD, 2006).

3.1.2.2 The Science and Society Approach

The scientific and technological revolutions at the end of the 20th century and the impacts of global environmental problems drew the attention to the values and the ethical and moral responsibilities of scientific education. Parallel to the science literacy concepts rooted in the products and processes of science, which is associated with the traditional school teaching of science (Roberts, 2007) other models emerged blending sciences with other disciplines (with social sciences, such as sociology) (Aikenhead, 2007). The science literacy concepts emphasizing science knowledge, skills and abilities were supplemented with the requirements of (1) scientific methods essential to understand other models of reality, (2) knowing and understanding the interaction between science, technology and society and (3) being responsible and environmentally aware (Klopfer, 1991; OECD, 2006; Riess, 2000).

Forming complex and plural science literacy is only possible by means of programs which lay an emphasis on the relationships between humans, nature and technology, such as Science Technology Society (STS) based projects. The fundamental principles of STS science are that the solutions for social problems and the satisfaction of needs/demand are supplied by scientific research and the direct executive is the technology. STS similarly to IBLE puts the individual into focus who can both form and apply knowledge and it transmits the socially relevant science knowledge embedded into a meaningful technological and social context (Aikenhead, 1994, 2000, 2003).

By integrating the experiences and philosophy of the STS programs into the IBSE methods the literacy which forms the basis of socially responsible decisions facilitating sustainability can be transmitted more efficiently. Understanding social issues with scientific background (e.g. energy production, environmental protection, greenhouse effect, increase in population), collecting information for their analysis by means of scientific inquiry and research; then discussing the results/conclusions and providing socially and technologically relevant feedback facilitates the direct and indirect transfer of science knowledge and its application in everyday situations. With the active exploration, scrutiny of the scientific background of ecological, social and economic problems by means of scientific methods and in the debates relying on and utilizing arguments, experiences students get to know and understand themselves as well; therefore, their motivation becomes more apparent. (Machamer, 1998). This in turn may form a good basis for the development of a responsible environmentally friendly attitude and behaviour.

3.1.3 Improving the Quality of Scientific Knowledge

3.1.3.1 General Goals, Integrated and Multidisciplinary Approaches

The demands and requirements regarding the acquisition of knowledge have shifted significantly over the past decade. The knowledge storage and reproduction have been overshadowed by meaningful learning and formulating a well-organized and efficiently applicable system of knowledge. A major role is attributed to the acquisition of knowledge taking into consideration the relationship between science and society based on scientific principles. “Science education has multiple goals. It should aim to develop understanding of a set of big ideas in science which include ideas of science and ideas about science and its role in society; scientific capabilities concerned with gathering and using evidence; scientific attitudes.” (Harlen, 2010, p. 8).

Social expectations and psychological research focusing on the organization of knowledge have made an impact on the aspects of selecting and organizing scientific content. In addition to the disciplinary perspective, which represents the systematic transmission of scientific knowledge following the logic of particular disciplines, inter- and multidisciplinary curriculum development emphasizing both the disciplinary integration and social aspects (Venville, Rennie & Wallace, 2009) also emerged. Therefore, teaching science does not only entail the understanding of basic concepts, themes, and experimental methods of particular disciplines. It is also imperative to include more comprehensive themes and cross-disciplinary content (e.g. the relationship between structure and function, the processes within and between systems, and the complex interdependencies of science, society and technology) (Klieme et al., 2003), ideas of science and ideas about science (Harlen, 2010, pp. 21–23) in science curriculum and standards.

To design the teaching process and to facilitate the acquisition of knowledge described by different curriculum and standards, it is important to take into consideration the findings of educational and psychological research regarding the organization and acquisition of knowledge (Duit & Treagust, 1998). Moreover, awareness about the difficulty and complexity of children's understanding of natural phenomena and the acquisition of scientific results and abstract conceptual systems, which is often accompanied by problems and flaws in interpretation also bears a special relevance. Misinterpretation of these processes and phenomena often requires students to reorganize their existing knowledge, modify their naive beliefs, and overwrite their misconceptions.

Meaningful learning is a prerequisite for understanding scientific knowledge and creating an efficiently applicable hierarchical conceptual system, which paves the way for students to be able to integrate new information into their existing knowledge by establishing meaningful connections between them (Ausubel, 1968; Roth, 1990). Matching existing and new knowledge – as highlighted by the constructivist approach of learning – requires the active cognitive efforts of students (Pope & Gilbert, 1983; Glaserfeld, 1995); its outcome is highly dependent on the quality of prior knowledge.

Children therefore do not start their formal education with a tabula rasa but already have their naive beliefs explaining the world around them. Learning can proceed smoothly if there is no contradiction between the experiential and the scientific knowledge, since this allows the easy assimilation of knowledge and the uninterrupted expansion of the conceptual system (e.g., the properties of living organisms). Misconceptions are likely to appear when experiential knowledge cannot be reconciled with scientifically-based theories. For example, children's Aristotelian worldview of body motion (motion must have a cause, in the absence of a causal factor, the body will be at rest) cannot be translated into the theoretical model of Newtonian mechanics (motion does not stop spontaneously, in an inertial reference frame bodies not subject to forces are either stationary or move in a straight line at a constant speed). Children may overcome the interpretational problem arising when learning Newtonian mechanics in several ways. They may form misconceptions by mixing the old and new knowledge and by distorting the new information to a lesser or greater extent, or they may memorize the new information without meaningfully assimilating it into their existing knowledge system. A common phenomenon is that children separate everyday experiences from the knowledge learnt at school, thus creating parallel explanations of the world, an everyday and a classroom knowledge base (Duit, 1994).

When the naive theory and the scientific knowledge are incompatible, substantial cognitive effort is required for learners to be able to understand and accept scientific knowledge. They are forced to revise their naive theories and restructure their prior knowledge and conceptual system. The difficulties students have to face as they reconcile their everyday beliefs with the scientific views are comparable to the paradigm shifts observed in the history of science as described by Kuhn (1962), like, for instance, the recognition of the heliocentric world view in place of the geocentric world view, or the replacement of the Newtonian theory with the theory of relativity (Arabatzis & Kindi, 2008).

A prerequisite to the abandonment of misconceptions or the prevention of their emergence is that students should be aware of their own beliefs and implicit assumptions about the world and compare their theories to the accounts given by their peers or by science. Opportunities to do so are provided by conversations, discussions and teacher or student experiments where students are given explanations for everyday phenomena. The process of shaping a conceptual system and evaluating one's own knowledge requires high cognitive engagement, reflectivity, metaconceptual awareness and advanced reasoning skills (Vosniadou & Ioannides, 1998). It is very important for students to realize that their beliefs are not facts but hypotheses that need to be tested, and that what they believe to be true has restricted validity and may turn out to be false in another system, in a different conceptual framework or at a different level of cognition (Vosniadou & Kollias, 2003). Inquiry-based science education may also contribute to a more conscious shaping of the conceptual system and facilitation of conceptual change.

The skills (questions, formulating hypotheses, research design, collecting, analyzing, evaluating and discussing data) developed during inquiry-based learning can be used in shaping the conceptual system and discussing different perspectives and interpretations. The tasks and problems used in inquiry-based learning provide an opportunity to take into consideration the various aspects of conceptual development: topics difficult to understand for students and requiring conceptual change can also be integrated. Inquiry-based learning can contribute to enhanced conceptual development, a more consistent concept system, fewer misconceptions, deeper understanding, and more meaningful scientific knowledge. The cooperative teaching methods applied in inquiry-based science education have a positive impact on the process of social knowledge construction and the development of social and communication skills.

3.1.3.2 Discipline-specific issues, preparing students for a science-related profession

IBSE may play a role in arousing interest in scientific issues and in students' acquisition of science literacy. In addition, it may also be useful in a more specialized and advanced science course for students who wish to embark on a scientific career. The reinterpretation of the role of enquiry and research appears more and more frequently in the methodology literature of science education. Erduran (2001) claims that in chemistry observation and experimenting are also an integral part of the traditional science education, but most importantly they are used to prove and confirm the facts included in the subject matter. We can speak of innovation in chemistry education if there are classroom activities directed at discovering the nature of the discipline of chemistry, the essence of the activity chemists are involved in, and if students learn to model the structure and function of matter.

This is in line with the fact that both science standards (National Research Council, 2000) and the PISA science literacy framework (OECD, 2006) highlight the importance of knowledge about the nature of science (NOS) and nature of scientific inquiry (NOSI). Whereas NOS refers to the product of inquiry, and science knowledge, NOSI refers to the processes of inquiry, the formation and acceptance of science knowledge. The main characteristic features of NOS were summarized by Schwartz, Lederman & Crawford (2004) as follows (Table 1).

Table 1 NOS Aspects and descriptions that served as a basis for comparison (Schwartz, Lederman & Crawford, 2004, p. 613)

Aspect	Description
Tentativeness	Scientific knowledge is subject to change with new observations and with the interpretations of existing observations. All other aspects of NOS provide rationale for the tentativeness of scientific knowledge.
Empirical basis	Scientific knowledge is based on and/or derived from observations of the natural world.
Subjectivity	Science is influenced and driven by the presently accepted scientific theories and laws. The development of questions, investigations, and interpretations of data are filtered through the lens of current theory. This is an unavoidable subjectivity thet allows science to progress and remain consistent, yet also contributes to change in science when previous evidence is examined from the perspective of new knowledge. Personal subjectivity is also unavoidable. Personal values, agendas, and prior experiences dictate what and how scientists conduct their work.
Creativity	Scientific knowledge is created from human imaginations and logical reasoning. This creation is based on observations and inferences of the natural world.
Sociocultural embeddedness	Science is a human endeavor and is influenced by the society and culture in which it is practiced. The values of the culture determine what and how science is conducted, interpreted, accepted, and utilized.
Observation and inference	Science is based on both observation and inference. Observations are gathered through human senses or extensions of those senses. Inferences are interpretations of those observations. Perspectives of current science and the scientist guide both observations and inferences. Multiple perspectives contribute to valid multiple interpretations of observations.
Laws and theories	Theories and laws are different kinds of scientific knowledge. Laws describe relationships, observed or perceived, of phenomena in nature. Theories are inferred explanations for natural phenomena and mechanisms for relationships among natural phenomena. Hypotheses in science may lead to either theories or laws with the accumulation of substantial supporting evidence and acceptance in the scientific community. Theories and laws do not progress into one and another, in the hierarchical sense, for they are distinctly and functionally different types of knowledge.
Interdependence of these aspects	None of these aspects can be considered apart from the others. For example, tentativeness of scientific knowledge stems from the creation of that knowledge through empirical observation and inference. Each of these acts is influenced by the culture and society in which the science is practiced as well as by the theoretical framework and personal subjectivity of the scientist. As new data are considered and existing data reconsidered, inferences (again made within a particular context) may lead to changes in existing scientific knowledge.

The characteristic features of NOSI are defined by Schwartz, Lederman & Lederman (2008, p. 4) as follows: "a) Questions guide investigation, b) multiple methods of scientific investigations, c) multiple purposes of scientific investigations, d) justification of scientific knowledge, e) recognition and handling anomalous data, f) sources, role, and distinctions between data and evidence, and g) community of practice". These features are considered to be general in science research and their understanding is facilitated by activities suggested by IBSE. However, it is also worth paying attention to the differences between the processes of understanding. Based on implications of disciplinary philosophy, Schwartz & Lederman (2008) highlight the importance of differences between scientific disciplines. Among others they refer to Spieker's (1972) work, who established an order between fields of natural science as follows: physics, chemistry, geology, botany and zoology. This order reflects an increase in the number of variables, a decrease in the application of mathematics and the fragmented nature of the discipline. Physics and chemistry "derive fundamental and universal laws from relatively small and concentrated bodies of data, or so to support generalizations reached through deductive reasoning" (1972, p. 75) whereas geology and biology typically require large and diverse amounts of data in order to make generalizations (see also Shayer, 1970).

3.2 Affective Outcomes

For a number of reasons cognitive goals have dominated education for centuries, including present science education. Among these reasons are the ways curricula are developed, considerations of the ease of explaining cognitive content, and the recent impact of cognitive science. However, research on the affective domain is becoming more influential. A number of affective areas, e.g. attitudes, attachment, beliefs, emotions, ethics, motivation, personal epistemologies, self-concept, self-regulation, social interaction etc. have recently been explored in relation to science education as well. Many policy documents include affective issues, national and international assessments pay growing attention to the assessment of noncognitive outcomes (Levin, 2013).

In this section, several factors associated with the affective domain of personality are described from the viewpoint of inquiry-based science education. These factors can be both the prerequisite constituents of students' achievement and the valuable outcomes of their learning (Bloom, 1977). We focus on the assessment of these factors as well as on their roles in inquiry-based science education.

3.2.1 Interest and Motivation

How does IBSE improve students' interest in science? And the other way round, how does students' interest improve performance via IBSE learning settings? To answer these questions, some general characteristics of the nature of students' interest in science learning need to be considered.

Students, by the very nature of their general interest in what researchers may call 'seductive details' (Harp & Mayer, 1998) become interested in science text learning – although, the positive effect of these entertaining and interesting pieces of information is questionable. Students' (and, of course, teachers') needs and interests are important factors in deciding what to teach. For example, in the topic of 'the nature of science' (NOS) teachers feel that it is the NOS part of the curriculum that conveys the valuable scientific thinking to students (Abd-El-Khalick, Bell & Lederman, 1998). However, teachers often find it difficult to teach NOS because "kids are not interested in the NOS itself" (p. 428). Hofstein and Rosenfeld (1996) claim that it is a wide repertoire of teaching strategies (including the involvement of out-of-school learning settings) that helps maintain students' interest in learning science.

According to the Rocard-report (Rocard et al., 2007, p. 3.), there is an "alarming decline in young people's interest" worldwide and especially in Europe towards science learning. According to this report, this decline can to a large extent be attributed to the way science is taught in schools.
The assessment of students' interest in learning science may have several forms. Open-ended questions about the nature of scientific inquiry (as used in Abd-El-Khalick, Bell and Lederman's, 1998, study) provide information about how pre-service teachers think about different aspects of scientific inquiry. Pre-service teachers often think that creativity plays a role in the design phase of experimentation, but they rarely think that data analysis requires creativity. The terms creativity and interest are intertwined in Renzulli's (1977) Interest-A-Lyzer questionnaire that helps students identify what they are really interested in by means of using open-ended and closed questions. Renzulli and Dai (2001) emphasized that interest and abilities can be studied in the context in which they manifest themselves. "We will never know whether a child is interested in astronomy or biology unless he or she is exposed to relevant materials" (p. 38). This quotation points to the two-sided nature of interests: both as antecedents and outcomes of the learning process.

The term interest is often cited paired with other terms: motivation and interest. The term motivation is even more often used in describing the psychological structures in mind that account for "why people think and behave as they do" (Graham & Weiner, 1996, p. 63). Järvelä and Niemivirta (1999) review different theories of motivation. From an inquiry-based education viewpoint we would like to emphasize the following line of thoughts: It is widely assumed that if the learning tasks are more interesting then this increases students' intrinsic motivation. The problem is that either new learning contexts require an already existing level of motivation or merely working on such 'interesting tasks' will result in task engagement. The debate here is to what extent motivation (especially intrinsic motivation) is an antecedent of successful learning or the outcome of being exposed to certain kinds of tasks and methods. One efficient way of improving students' motivation in high school science is the use of digital games. According to Papastergiou's (2009) results, this tool is not only motivating, but also effective both for boys and girls.

For the assessment of motivation, the paper-and-pencil questionnaire is by far the most widespread method. (Most of these questionnaires can be equally well used as on-line tests administered via ICT tools.) One example of a motivation questionnaire for science learning contains six scales measuring different aspects of the affective components of learning. One of the 35 items explicitly addressed inquiry activities in learning: "In science, I think, it is important to participate in inquiry activities". The items used five-point Likert-scale from 'strongly disagree' to 'strongly agree'. The six scales are listed below (Tuan, Chin & Shieh, 2005, p. 643).

1. Self-efficacy. Students believe in their own ability to perform well in science learning tasks.
2. Active learning strategies. Students take an active role in using a variety of strategies to construct new knowledge based on their previous understanding.
3. Science learning value. The value of science learning is to let students acquire problem-solving competency, experience the inquiry activity, stimulate their own thinking, and find the relevance of science with daily life. If they can perceive these important values, they will be motivated to learn science.
4. Performance goal. The student's goals in science learning are to compete with other students and get attention from the teacher.
5. Achievement goal. Students feel satisfaction as they increase their competence and achievement during science learning.
6. Learning environment stimulation. In the class, learning environment surrounding students, such as curriculum, teachers' teaching, and pupil interaction influenced students' motivation in science learning.

Other methods have also been applied to explore students' interest in science. For example, students' self-generated questions may indicate how they see scientific activities and what areas of science are most well known or interesting for them (see e.g. Cakmakci, Sevindik, Pektas, Uysal, Kole & Kavak, 2012).

A third concept often tied together with interest and motivation is anxiety. There are clear connections between assessment methods and the level of anxiety. Any efforts that aim to improve teachers' formative assessment practice (e.g., Sato, Wei & Darling-Hammond, 2008) may yield positive consequences in students' anxiety reduction. This latter change further may increase students' motivation, since test anxiety is the main source of negative motivation (Hill & Wigfield, 1984).

3.2.2 Attitudes and Beliefs

Perhaps the most intensively investigated psychological constructs of the affective domain are attitudes towards fields and objects of study and beliefs about different phenomena connected to learning. Attitudes have been studied in the context of science education for a long time, they are involved in most large-scale national and international assessment program (for en extensive review of early research on science-related attitudes, see Gardner, 1975), while beliefs have recently become popular field of research. To differentiate between attitudes and beliefs, we follow McLeod's insight (see Andrews, Diego-Mantecon, Vankuš, Op 't Eynde & Conway, 2007) that beliefs, attitudes and emotions follow each other on a continuum indicating an increasing level of affective, and decreasing level of cognitive involvement. Therefore, measuring attitudes is usually accomplished by means measuring simple, mainly emotion-driven decisions about to what extent the student likes or dislikes something. According to a large-scale Hungarian study (Csapó, 2000), mathematics and science – attitude measured on a five-point Likert scale – gradually become less and less popular during compulsory education. In this study, no sex differences were found in the attitude towards mathematics and chemistry, but girls preferred biology, and boys were more in favour of physics both in grades 7 and 11. According to Csíkos' (2011) results among 7th grade students, biology was one of the most frequently named favourite school subjects, whereas geography, mathematics and physics were the least preferred subjects.

The importance of investigating students' beliefs has been highlighted in several recent studies from different aspects. One important aspect is the role of general epistemological beliefs in school achievement as revealed among others by Schommer (1993), and Muis and Franco (2009). Hofer (2002, p. 4.) provided a list of phenomena being in the focus of concern: "the definition of knowledge, how knowledge is constructed, how knowledge is evaluated, where knowledge resides, and how knowing occurs". In the field of mathematical beliefs, there is an agreement on the multidimensional characteristic of mathematical beliefs (De Corte, Op 't Eynde & Verschaffel, 2002), and in fact there have been several different factors found in empirical research (e.g., Andrews, Diego-Mantecon, Vankuš, Op 't Eynde & Conway, 2007). Limón (2006, p. 23.) emphasizes that "individuals may sustain different epistemological beliefs when [epistemological beliefs] are applied to different domains". The problem of domain-specificity of epistemological beliefs is to a great extent the problem of methodology.

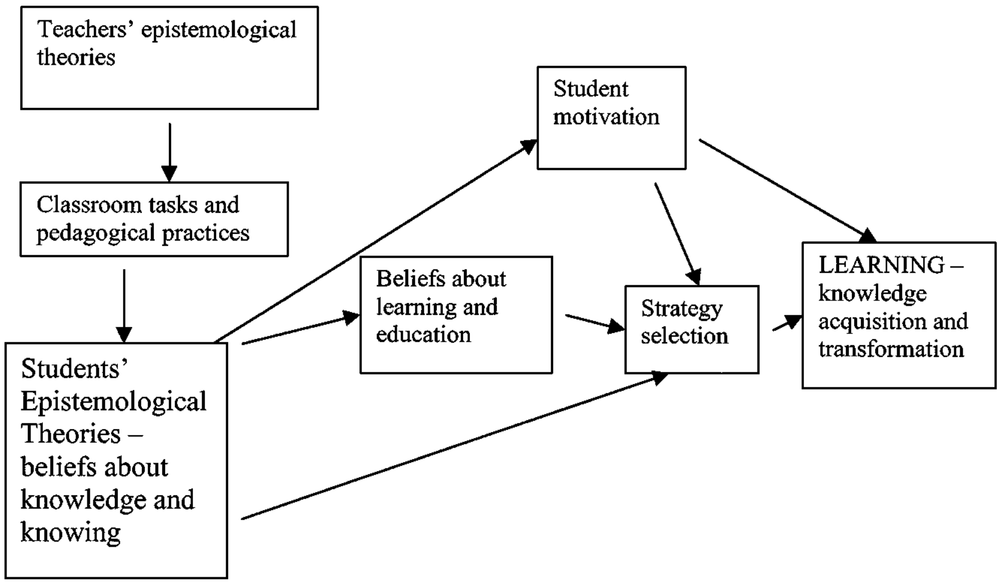

Figure 1 Working model of how epistemological theories influence classroom learning. (Hofer, 2001 p. 372)

Bell and Linn (2002) reviewed different types of measures of beliefs about science. They argue that it is very difficult to change students' understanding of scientific inquiry. One of the tasks that indicate such difficulties is the following (Bell and Linn, 2002, p. 329):

> Three rods (plastic, diamond and copper) are doused into boiling water. After 10 minutes the temperature at the other end of each rod is measured. The diamond rod was the hottest, and the plastic was the coolest. Can it be concluded that diamond is the best conductor of heat? Why or why not?

Bell and Linn (2002) analyzed the views lay people hold about science and scientific inquiry. These views are often inconsistent: students who critique advertisements can later use the same argument (cite-amnesia). The serendipitous, personality-filled and controversial nature of most scientific breakthroughs is often hidden in both textbooks and in journal articles. The Internet may provide an opportunity for developing critical evaluation of scientific information, and careful reflection on different sources of information. The importance of reflecting to our own scientific ideas and beliefs is emphasized by Elder (2002) who suggests that even elementary school students can be receptive to those forms of instruction which advance their epistemological beliefs. Students' epistemological beliefs can be measured by means of several (mostly paper-and-pencil based) questionnaires; however, the reliability of these instruments is often problematic.

The importance of measuring students' beliefs about learning science resulted in very detailed background questionnaires in international system-level surveys. Both TIMSS and PISA use detailed questionnaires containing several items about students' classroom practice and beliefs. In the TIMSS science questionnaire for Grade 8 students, students judged different statements about the school lessons and about their own beliefs on 4-point Likert scale. This type of Likert-scales disables the option of choosing the neutral position, and students are compelled to choose either agreement or disagreement. For example, the difficulty and the boredom of a subject, and some statements about the importance and future-orientations were administered to students.

The questionnaire for PISA 2006 (when science was the domain in focus) also builds on 4-point Likert-scales. The topics covered include: items on enjoyment and interest; easiness of different knowledge and knowledge acquisition forms; societal usefulness of science; frequency of using different information sources when learning science; interest in and responsibility for environmental issues; the sources of information about environmental issues etc. From a methodological point of view, the huge amount of data available (in part due to the Likert-scale format) makes the PISA studies a valuable information source and a reference point when interpreting other empirical results. How students think about environmental issues can be both the antecedents and the result of how science is taught. Inquiry-based science education has the chance to shape and develop students' views about science learning in general and about the most important (often global) environmental issues.

The possible implications of PISA results from the viewpoint of real world contexts have been discussed by Fensham (2009). The results suggest that countries have consistently relatively high (or low) scores on these so-called 'contextual set of items'; therefore it was not the concrete context that may have caused difficulty, but the real world context in general.

3.2.3 Self-Concept and Future-Orientation

Traditional education methods quite often emphasize declarative knowledge and present science as a set of information to be mastered. Students may develop a false self-concept, as they believe they may not be good in science, or those may be successful who are good at rote learning. It is plausible that students with false academic self-concept will not fully realize their academic achievement potential. There is empirical evidence of the reciprocal relationship between self-concept and achievement among elementary school children (Guay, Marsh & Boivin, 2003). Guay, Ratelle, Roy and Litalien (2010) revealed that among high school students it is academic motivation that mediates the effect between academic self-concept and achievement.

Investigating academic self-concept usually takes the form of questionnaires with Likert-scale items. For example, Boivin, Vitaro and Gagnon (1992) developed a 36 item scale called Self-Perception Profile for Children comprising 4-point Likert scale items. This questionnaire used an item format that helps avoiding 'good child' oriented (socially desirable) answers, i.e. intended to measure the real self-concept instead of an ideal one.

The Rocard-report (Rocard et al, 2007) analyzes the phenomenon of girls' lower interest and participation in science learning, and claims that the "pattern of gender differences continues with women choosing fewer academic studies in math, science and technology (MST). In fact, at the European level, girls account only for 31% of MST graduates" (p. 6). Girls' career choices are affected by several factors. According to Eccles (1994, p. 604), "women are less likely to enter these fields [mathematics and physical science] than men, both because they have less confidence in their abilities and because they place less subjective values on these fields". Education may find its role in shaping both factors underlined in Eccles' model.

As cited earlier, the Rocard-report states that the decline in career choices in the fields of mathematics, science and technology can be to a large extent attributed to the way how science is taught in schools. Therefore this high level expert group proposed that research and dissemination programs aiming to promote inquiry-based learning and teaching across Europe be supported. SAILS among other projects may contribute to the Europe-towards-2020 aims in that through IBSE approach much fewer students will be underachievers in the PISA science studies.

4 Assessments in the Context of the SAILS Project

Educational assessment is a well defined field of research and practice which deals with collecting, analyzing and utilizing data on students' learning outcomes (Black, 2000). A great variety of methods and instruments are available for educational assessment and measurement, and most of these are appropriate in the context of IBSE as well. Diagnostic assessment frameworks are usually created to help item and test development, and assessments by tests may be utilized to develop teachers' assessment skills and to help them make more objective decisions about students' knowledge, even when they use personal, face-to-face methods (Gulikers, Biemans, Wesselink & van der Wel, 2013). Non-diagnostic assessments are elaborated on in Deliverable 3.1; they concern chiefly the classroom assessments that go on during the teaching and learning process.

4.1.1 Formative and Summative Assessment

Educational assessment is a well defined field of research and practice which deals with collecting, analyzing and utilizing data on students' learning outcomes (Black, 2000). There are several definitions of educational assessment, but most of them share some core elements. The Joint Committee on Standards for Educational Evaluation (2003) offers this definition:

> "Assessment: The process of collecting information about a student to aid in decision making about the student's progress and development." (p. 5).

Definitions and interpretations of assessment usually distinguish two main purposes of the assessment. The characteristics mentioned by Harlen and Deakin Crick (2002) appear in most descriptions:

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

Formative assessment is embedded in teaching and learning and its main purpose is to improve students' learning. The focus therefore is on future learning and the means of getting to that next stage. It is often followed by an intervention which intends to compensate for developmental deficiencies. *Summative assessment* takes place at the end of a longer learning process and typically concerns itself with larger units of learning outcomes.

4.1.1.1 Formative Assessment

Formative assessment is sometime also referred to as assessment for learning (Black, Harrison, Lee, Marshall & Wiliam, 2003). The Joint Committee on Standards for Educational Evaluation (2003) defines formative evaluation as "Evaluation conducted while a creative process is under way, designed and used to promote growth and improvement in a student's performance or in a program development" (p. 228).

A number of recent books and papers describe the particular characteristics of formative assessment (e.g. Suskie, 2009; Sindler, 2011), assessment taking place in classroom environment (Kubiszyn & Borich, 2010, McMillan, 2013). Several publications them focus explicitly on formative assessment in science (Hodgson & Pyle, 2010; Hammerman, 2009; Coffey, Hammer, Levin & Grant, 2011; Hickey, Taasoobshirazi & Cross, 2012; Ruiz-Primo, Li, Wills, Giamellaro, Ming-Chih, Mason & Sands, 2012).

In the past decades, one direction of research on formative assessment aimed at establishing a strong theoretical grounding, and placing it in a broader context of educational evaluation (Black & Wiliam, 1998a, 2003, 2009; Black, Harrison, Lee, Marshall & Wiliam, 2002; Black, 2013). Researchers agree that defining characteristics of formative assessment are that it takes place during the learning process and not after it; it provides an immediate and detailed feedback for the students and/or teachers; and the assessment information is used to modify the learning process to make it more effective. Feedback is an especially important attribute of every assessment, and in formative assessment, time between collecting information on learning outcome and informing learners has to be short to maintain and improve learning process (Wiliam, 2013; Havnes, Smith, Dysthe & Ludvigsen, 2012).

Studies dealing with the practical aspects of non-diagnostic formative assessment describe a variety of ways information is gathered, the results are analyzed, interpreted and communicated to students (Black & Wiliam, 1998b; Black, Harrison, Lee, Marshall & Wiliam, 2003). Non-diagnostic formative assessment requires traditional classroom activities to be reorganized as the emphasis is on creating opportunities within the learning experience for both teachers and learners to review and challenge understanding and so requires a more dialogic approach. We know from previous research on the King's-Medway-Oxford–Formative assessment Project (Black et al, 2003) that: "many teachers do not plan and conduct classroom dialogue in ways that might help students learn. Put simply, the only point of asking questions is to raise issues about which teachers need information or about which the students need to think."

When interpreting formative assessment in the context of IBSE, we have to take into account that inquiry methods consider learning as an active and constructive process, thus the immediate feedback formative assessment provides should be integrated into this active learning process. Thus, formative assessment has to deal with components of knowledge where changes are observable after relatively short periods. In such situations, there may be more direct correspondence between classroom activities and learning outcomes, and the feedback the assessment provides may orient the next phase of learning.

Several previous studies have discussed issues especially relevant for SAILS. Research on teaching thinking has been developing rapidly since the 1980s, and in this paradigm assessment practices supporting the development of thinking have also been examined. For example, Stiggins, Griswold and Wikelund (1989) studied teachers' assessment of higher order thinking skills in their classroom work. The subjects included mathematics, science, social studies and language arts, where the assessment of five types of thinking (recall, analysis, comparison, inference, and evaluation) was examined. Several assessment acts and forms were observed and analyzed, such as teachers asking oral questions (5221 oral questions were recorded), teacher-made paper and pencil tests, text-embedded tests, and written assignments (altogether 4120 exercises were analyzed). The results indicated that an overwhelming proportion of questions were related to recall (65%), and only 11% involved analysis, 5% comparison, 17% inference, and 2% evaluation (Stiggins, Griswold and Wikelund (1989, p. 240, Table 2).

These statistics indicate that IBSE skills are typically poorly addressed by teachers' questioning. In diagnostics formative assessment, the Evidence-Based Reasoning Assessment System (EBRAS) can be used to reveal students' misconceptions and logical errors (Brown, Nagashima, Fu, Timms, & Wilson, 2010). This system describes assessments as a cyclical process which involves four steps:

> "(a) observing: using tasks to elicit performances assumed to depend upon the latent variables;
> (b) scoring: using rubrics to categorize different observed performances and assign them relative value;
> (c) summarizing: using a measurement model to aggregate the individual scores into measures of the latent variables; and

(d) interpreting: using the model of cognition to give meaning to the estimated values of the latent variables and, in so doing, answer the assessment question." (Brown, et al., 2010, p. 144).

Development of assessment items in this model is based on construct maps, which also help interpret the results. For example, the "conceptual sophistication construct map" identifies seven levels: unproductive misconception, productive misconception, singular, relational, combined, multi-relational and multi-combined (Brown, at al., 2010, p. 148, Fig. 7). The item development process of EBRAS is especially useful for formative assessment, as formative assessments are often composed of a single items, and an item developed in the EBRAS framework is placed in a complex system, thus students' actual developmental level can be better identified.

In a more recent paper, Heritage (2013) describes how data can be collected about student understanding. Interaction is identified as a primary source of evidence about understanding: first of all, the interactions between students and teacher, but further sources such as students' writings, drawings and other artifacts can also be analyzed.

In non-diagnostic formative assessment, teachers need to use a variety of tools to find where students are in their learning. From these data, they can make judgements that can help the student to decide on the next step in learning, and so guide them towards improvement. For this kind of feedback to function in a formative way, there are a number of prerequisites (Harrison, 2009):

- a need for teachers to create regular opportunities in the classroom for students to discuss and communicate their perception of their evolving learning;
- a willingness by teachers to develop or adapt future learning activities in response to learning needs and development;
- the capability of teachers to give and model descriptive feedback that encourages learners to make improvements to their work;
- an acceptance that learners need to be involved in decisions about their learning and are helped to develop the skills to do this;
- an awareness of the skills, ideas and concepts needed to produce quality pieces of work that recognises misconceptions, likely reasoning errors and mistakes as the beginning of developing better understanding.

The view of assessment developed by Black et al. (2003) and explored more theoretically in Black & Wiliam (2009) argues that it is not an extra feature of pedagogy, but rather that it should be an integral part of any model of pedagogy. Thus, if teachers start with the aim of promoting inquiry, their first step should be to design classroom activities which have the potential to involve students in dialogue, with the teacher and with one another, an involvement which can develop their capacity to engage as active learners. This central importance of dialogue is stressed by Alexander (2008) :

> "Children, we now know, need to talk, and to experience a rich diet of spoken language, in order to think and to learn. Reading, writing and number may be acknowledged curriculum 'basics', but talk is arguably the true foundation of learning." (p.9)

This point is developed further in Wood's (1998) reference to Vygotsky:

> "Vygotsky, as we have already seen, argues that such external and social activities are gradually internalized by the child as he comes to regulate his own internal activity. Such encounters are the source of experiences which eventually create the 'inner dialogues' that form the process of mental self-regulation. Viewed in this way, learning is taking place on at least two levels: the child is learning about the task,

> developing 'local expertise'; and he is also learning how to structure his own learning and reasoning." (p.98)

In this view of assessment, the potential of any planned activity has then to be realized by the way in which the teacher implements it in the classroom. The teachers' task requires great skill and judgment, both to encourage all students to participate, to promote peer discussion of alternative views, and yet to steer the discussion in a productive direction (Black & Wiliam, 2009). A subsequent step will naturally be some form of informal summative review, to check what has been achieved, either in the form of individual writing tasks or in informal tests of a specific topic.

It is the view of the SAILS project that both diagnostic and non-diagnostic formative assessment have a lot to offer to teachers who employ IBSE strategies.

4.1.1.2 Summative Assessment

Summative assessment may be relevant in the context of IBSE projects in several ways, as it is applied both in the context of classroom learning (Biggs, 1998; Harlen, 2013) and in large-scale assessments as well. Summative tests are used when the general efficiency of IBSE methods is studied. In such external evaluation of the inquiry learning, the tests applied are independent of the specific goals of IBSE. When IBSE and other approaches are compared, external criteria of efficiency are considered and the assessment instruments are based on principles different from inquiry learning. For example, Gee and Wong (2012) examined whether IBSE applied in the participating countries had a measurable effect on the countries' PISA results. Many experiments exploring the efficiency of IBSE compared it to other teaching methods, most often to an existing mainstream science education method, and summative tests of assessing science knowledge and skills were explored.

Based on the general goals associated with IBSE, several hypotheses may be constructed which then could be explored in teaching experiments using summative tests. For instance, if the hypothesis that a particular IBSE implementation improves problem solving is examined, problem solving tests should be employed to measure the development of problem solving skills of students taught by IBSE and non-IBSE methods. Many available summative tests created for assessments outside of IBSE projects may be used for such external evaluation of IBSE. If the general efficiency of IBSE is to be explored, tests based on TIMSS and PISA science frameworks and similar science tests can be used. If the impact of an IBSE project on skills specifically associated with IBSE is to be examined, the existing tests for the assessments of these skills may be used. However, comparing IBSE and non-IBSE methods requires a careful experimental design and well-controlled data collection conditions. Conducting such experiments is beyond the SAILS, but the assessment frameworks developed in the project may contribute to the theoretical grounding of such experiments.

For the internal evaluation of IBSE projects, summative tests based on inquiry learning goals may be used. At the end of a semester or a school year, summative tests may be used to find out whether the goals are met. These tests can be devised using the taxonomy described in the previous sections and the frameworks to be prepared in the next phases of the SAILS project.

4.1.2 Social Context of the Assessment

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

One of the major differences between large-scale testing and classroom assessment is that classroom assessment is more personal, takes place in a social context, and often involves interpersonal communication. Because of this personal nature, validity (Bonner, 2013) and reliability (Parkes, 2013) of classroom assessment requires special attention. Developing teachers' assessment competencies may be one of the main avenues to improve quality, as it is conceptualized in the SAILS project (see also Campbell, 2013). Combining formative assessment based on non-diagnostic assessments carried out as learning is taking place and diagnostic assessments to measure attainment before and after instruction is designed to improve the quality of both students' and teachers' assessment competencies.

4.1.2.1 Students' Self-Assessment

The first persons who can assess students' learning outcomes are the learners themselves. Their time can be allocated for the assessment activities more easily than teachers' time, and they are also motivated to find out their results as quickly as possible. On the other hand, students' judgment of their own performance may be rather biased, and, as untrained assessors, they may make errors most probably in the assessment process. Despite these constrains, students' self assessment may be potentially very useful and also important taking into account the requirements of life-long-learning: Students are expected to become independent learners being able to manage the entire learning process, including assessment. As IBSE offers opportunities for students' individual work, it involves numerous possibilities to develop self-regulated learning strategies and metacognition as well. Formative assessment also attempts to develop students' self-regulated learning strategies (Black, McCormick, James & Pedder, 2006; Clark, 2012), and integration of self-assessment may be promising for achieving these goals.

In IBSE projects, self-assessment may take place in various forms, e.g. students may report and evaluate their successes and difficulties in the inquiry processes (Topping, 2003). Reviewing assessment practices in science education almost two decades ago, Tamir (1996), found only a few instances of self assessment in the context of science. Among those, the most detailed one was the Self-Report Knowledge Inventory (SRKI), which helped students evaluate their perceived level of mastery on a five-point scale. In a more recent review Brown and Harris (2013) found far more studies dealing with self-assessment. In their systematic literature review they describe methods of students' self-assessment as they "focus directly on obtaining from students an estimate or description of how well they believe they will do or have done on a specific test or task". They classify self-assessment practices into three major groups: "(1) self ratings, (2) self-estimates of performance, and (3) criteria- or rubric-based assessments" (Brown & Harris, 2013. p. 369). One of the conclusions of their review is that training in self-assessment strategies improves learning and results in measurable gain. The use of self-assessment produced more remarkable effects when self-regulated learning strategies were systematically taught, and when self-assessment was accompanied by other types of assessment. In IBSE, each form of self-assessment may be relevant, and the related competencies may be systematically taught by supporting students in assessing their own inquiry activities.

4.1.2.2 Student-Student Communication and Peer Assessment

Collaboration and teamwork is a typical setting of activities in modern societies, and assessing the related competencies has received growing attention. For example, in PISA 2015, collaborative problem solving will be the innovative assessment domain. Science inquiries may be carried out individually or in groups, and in both cases there are a number of contexts where competencies needed for collaborative activities may be fostered. Similarly, there are several opportunities where students may evaluate each-other's learning, and give useful feedback to their peers. Using peer assessment may improve important social skills, communication skills and collaborative skills (Topping, 2003).

References

Abdellatif, H. R., Cummings, R. & Maddux, C. D. (2008). Factors affecting the development of analogical reasoning in young children: a review of literature. *Education*, 129(2), 239-249.

Abd-El-Khalick, F., Bell, R. L. & Lederman, N. G. (1998). The nature of science and instructional practice: making the unnatural natural. *Science Education*, *82*, 417-436.

Adey, P. (1992). The CASE results: implications for science teaching. *International Journal of Science Education*, 14(2), 137-146. doi: 10.1080/0950069920140202

Adey, P. (1999). Thinking Science: Science as a gateway to general thinking ability. In: Hamers, J. M., Van Luit, J. E. & Csapó, B. (Eds), *Teaching and learning thinking skills*, (pp. 63-80). Lisse: Swets and Zeitlinger.

Adey, P., Csapó, B., Demteriou, A., Hautamäki, J. & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, 2(2), 75-97.

Adey, P., Nagy, F., Robertson, A., Serret, N. & Wadsworth, P. (2003). *Let's think through science!* London: nferNelson.

Adey, P. & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention programme in the high school science curriculum. *Cognition and Instruction*, 11(1), 1-29.

Adey, P. & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.

Adey, P., Shayer, M. & Yates, C. (2001). *Thinking Science: The curriculum materials of the CASE project* (3rd ed.). London: Nelson Thornes.

Aikenhead, G. S. (2007). Expanding the research agenda for scientific literacy. Paper presented to the "Promoting Scientific Literacy: Science Education Research in Transaction" Uppsala University, Uppsala, Sweden, 28-29 May 2007.

Aikenhead, G., S. (1994). What is STS teaching? In J. Solomon & G. S. Aikenhead (Eds.). *STS education: International perspectives on reform* (pp. 47-59). New York: Teachers College Press.

Aikenhead, G., S. (2000). STS in Canada: From policy to student evaluation. In D. D. Kumar & D. E. Chubin (Eds.), S*cience, technology and society. A sourcebook on research and practice*. New York: Kulwer Academic/Pleum Publishers. http://www.usask.ca/education/people/aikenhead/

Aikenhead, G., S. (2003). STS Education: A rose by any other name. In T. Cross (Ed.), *A vision for science education: Responding to the work of Peter J. Fensham* (pp. 59–75). London: Routledge. http://www.usask.ca/education/people/aikenhead/

Ainsworth, L. & Viegut, D. (2006). *Common formative assessments. How to connect standards-based* instruction *and assessment.* Thousand Oaks, CA: Corwin Press.

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

Alfieri, L., Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1-18.

Alexander, R. (2008). *Towards dialogic thinking: rethinking classroom talk.* (4th Ed.) York, England: Dialogos.

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E. & Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. *Cognitive Development*, 23(4), 452-471.

Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing.* New York: Longman.

Anderson, R. D., (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13,* 1–12.

Andrews, P., Diego-Mantecon, J., Vankuš, P., Op‘t Eynde, P. & Conway, P. (2007). Construct consistency in the assessment of students’ mathematics-related beliefs: a four-nation, cross-sectional factor-analytic study. Paper presented at the 12th Biennial Conference for Research on Learning and Instruction, Budapest, Hungary.

Arabatzis, T. & Kindi, V. (2008). The problem of conceptual change in the philosophy and history of science. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 345–373). New York: Routledge.

Ausubel, D. P. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston.

Bell, P. & Linn, M. C. (2002). Beliefs about science: How does science instruction contribute? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 321-346). Mahwah, NJ: Erlbaum.

Bennett, R. E., Persky, H., Weiss, A. R. & Jenkins, F. (2007). *Problem solving in technology-rich environments: A report from the NAEP Technology-Based Assessment Project* (NCES 2007–466). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Bernholt, S., Neumann, K. & Nentwig, P. (Eds.) (2012). *Making it tangible – Learning outcomes in science education*. Münster: Waxmann.

Bird, L. (2010). Logical reasoning ability and student performance in general chemistry. *Journal of Chemical Education*, 87(5) 541-546.

Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education*, 26(3-4), 407-419. doi: 10.1080/713688540

Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 167-178). London: SAGE.

Black, P. & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: The science student’s role in formative assessment. *School Science Review*, 83(302), 43-49.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2002). Working inside the black box: assessment for learning in the classroom. London: King’s College London, Department of Education and Professional Studies.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning. Putting it into practice.* Berkshire: Open University Press.

Black, P., McCormick, R., James, M. & Pedder, D. (2006). Learning how to learn and assessment for learning: a theoretical inquiry. Research Papers in Education. 1(2) 119-132.

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability,* 21(1), 5-31. doi: 10.1007/s11092-008-9068-5

Black, P. & Wiliam, D. (2003). ‘In praise of educational research’: formative assessment. *British Educational Research Journal*, 29(5), 623-637. doi: 10.1080/0141192032000133721

Black, P. & Wiliam, D. (1998a) *Inside the black box: raising standards through classroom assessment*. London: King’s College London School of Education.

Black, P. & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice,* 5(1), 7-73. doi: 10.1080/0969595980050102

Bloom, B. S. (1977). Affective outcomes of school learning. *The Phi Delta Kappan*, 59, 193-198.

Boivin, M., Vitaro, F. & Gagnon, C. (1992). A reassessment of the Self-Perception Profile for Children: Factor structure, reliability, and convergent validity of a French version among second through sixth grade students. *International Journal of Behavioral Development,* 15, 275-290. DOI: 10.1177/016502549201500207

Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 87-106). London: SAGE.

Boyera, T. W., Levinea, S. C. & Huttenlochera, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, 44(5), 1478-1490.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes. A quantitative synthesis. *Review of Educational Research,* 53, 499-518. doi: 10.3102/00346543053004499

Brown, G. T. L. & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 87-106). London: SAGE.

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M. & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, 15(3-4), 142-174. doi: 10.1080/10627197.2010.530562

Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 71-84). London: SAGE.

Coffey, J. E., Hammer, D., Levin, D. M. & Grant, T. (2011). The missing disciplinary substance of formative assessment. Journal of Research in Science Teaching. 48(10) 1109-1136. doi: 10.1002/tea.20440

Cakmakci, G., Sevindik, H., Pektas, M., Uysal, A., Kole F.,& Kavak, G. (2012). Investigating Turkish primary school students’ interest in science by using their self-generated questions. *Research in Science Education*. 42(3), 469-489.

Caroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3). www.amstat.org/publications/jse/v10n3/chance.html

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. Educational Psychology Review. 24(2) 205-249. doi: 10.1007/s10648-011-9191-6

Comber, L. C. & Keeves, J. P. (1973). *Science education in nineteen countries. International Studies in Evaluation.* New York: Wiley.

Csapó, B. (1992). Improving operational abilities in children. In A. Demetriou,. M. Shayer & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education* (pp. 144–159). London: Routledge and Kegan.

Csapó, B. (1997). Development of inductive reasoning: Cross-sectional measurements in an educational context. *International Journal of Behavioral Development*, 20(4), 609–626.

Csapó, B. (1999). Improving thinking through the content of teaching. In: J. H. M. Hamers, J. E. H. van Luit & B. Csapó (Eds.), *Teaching and learning thinking skills* (pp. 37-62). Lisse: Swets and Zeitlinger.

Csapó, B. (2000). A tantárgyakkal kapcsolatos attitűdök összefüggései [Connections between school subject attitudes] *Magyar Pedagógia, 100*, 343-366.

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

Csapó, B., Ainley, J., Bennett, R., Latour, T. & Law, N. (2012). Technological issues of computer-based assessment of 21st century skills. In B. McGaw & P. Griffin (Eds). *Assessment and teaching of 21st century skills.* New York: Springer. 143-230.

Csapó, B., Molnár, Gy. & R. Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills. A pilot study for introducing electronic testing in large-scale assessment in Hungary. In F. Scheuermann & J. Björnsson (Eds.). *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 120-125). Luxemburg: Office for Official Publications of the European Communities.

Csapó, B. & Szabó, G. (Eds.) (2012). Framework for diagnostic assessment of science. Budapest: Nemzeti Tankönyvkiadó.

Csapó, B. & Szendrei, M. (Eds.). (2011). *Framework for diagnostic assessment of mathematics*. Budapest: Nemzeti Tankönyvkiadó.

Csíkos, C. (2011). The place for mathematics in students' beliefs about learning. In B. Rösken & M. Casper (Eds.), *Current state of research on mathematical beliefs XVII. Proceedings of the MAVI-17 Conference* (pp. 43-53). Bochum: Professional School of Education, Ruhr-Universität Bochum.

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., Howe, A. (2013). Creative learning environments in education – A systematic literature review. *Thinking Skills and Creativity,* 8, 80-91.

De Corte, E., Op 't Eynde, P. & Verschaffel, L. (2002). ''Knowing what to believe'': The relevance of students' mathematical beliefs for mathematics education. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 297–320). Mahwah, NJ: Erlbaum.

DeBoer, G. E. (1991). A History of Ideas in Science Education: Implications for Practice. New York: Teachers College Press.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37(6). 582-601.

Duit, R. & Treagust, T. F. (1998). Learning in science – From behaviourism towards social constructivism and beyond. In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education* (pp. 3–25). Boston: Kluwer Academic Publishers.

Duit, R. (1994). Research on students' conceptions – developments and trends. In H. Pfundt & R. Duit (Eds.), *Bibliography: Students' alternative frameworks and science education* (pp. 22–42). Kiel: Institute for Science Education at the University of Kiel.

Durant, J. (1994). What is scientific literacy? *European Review*, 2, 83-89.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. Applying the Eccles et al. Model of Achievement-Related Choices. *Psychology of Women Quarterly,* 18, 585-609.

Education Council and European Commission (2006). Modernizing education and training: a vital contribution to prosperity and social cohesion in Europe joint interim report of the council and of the commission on progress under the 'education & training 2010' work programme. Brussels: Education Council and European Commission.

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

Elder, A. D. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 347-363). Mahwah, NJ: Erlbaum.

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

Erduran, S. (2001). Philosophy of chemistry. In F. Bevilacqua, E. Giannetto & M. R. Matthews (Eds.), *Science education and culture. The contribution of history and philosophy of science* (pp. 165– 176). Dordrecht: Kluwer Academic Publishers.

Fensham, P. J. (2009). Real world contexts in PISA science: implications for context-based science education. *Journal of Research in Science Teaching*, 46, 884-896.

Fischer, A., Greiff, S. & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, *4*, 19–41.

Frensch, P. A. & Funke, J. (Eds.). (1985). *Complex problem solving: The European perspective.* Hillsdale, NJ: Lawrence Erlbaum.

Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, *11*, 133–142. doi:10.1007/s10339-009-0345-0

Furtak, E. M., Seidel, T., Iverson H. & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82(3), 300-329. doi: 10.3102/0034654312457206

Gardner, P. L. (1975). Attitudes to science: A review. *Studies in Science Education.* 2(1), 1-41.

Gee, K., A. & Wong, K. K. (2012). A cross national examination of inquiry and its relationship to student performance in science: Evidence from the Program for International Student Assessment (PISA) 2006. *International Journal of Educational Research*, 53, 303-318. doi: 10.1016/j.ijer.2012.04.004

Girotto, V. & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition,* 106(1), 325-344.

Glaserfeld, E. (1995). *Radical constructivism. A way of knowing and learning.* London, Washington, DC: The Palmer Press.

Graham, S. & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 63-84). New York: Macmillan.

Greiff, S., Wüstenberg, S. & Funke, J. (2012). Dynamic Problem Solving: A new assessment perspective. *Applied Psychological Measurement*, *36*, 189–213. doi:10.1177/0146621612439620

Greiff, S., Wüstenberg, S., Molnár, Gy., Fischer, A., Funke. J. és Csapó, B. (2013): Complex Problem Solving in Educational Contexts – Something beyond g: Concept, Assessment, Measurement Invariance, and Construct Validity. *Journal of Educational Psychology,* 105(2), 364-379. doi: 10.1037/a0031856.

Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95, 124-136.

Guay, F., Ratelle, C. F., Roy, A. & Litalien, D. (2010). Academic self-concept, academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, 20, 644-653.

Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R. & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation*, 39(2), 116-124. doi:10.1016/j.stueduc.2013.03.001

Hackling, M. W. & Prain, V. (2008). Research Report 15: Impact of primary connections on students' science processes, literacies of science and attitudes towards science. Canberra: Australian Academy of Science. http://www.science.org.au/primaryconnections/irr-15.pdf

Hambleton, R., Swaminathan, H. & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California: SAGE.

Hamers, J. H. M., de Koning, E. & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, 23(2), 132-148.

Hammerman, E. (2009). *Formative assessment strategies for enhanced learning in science, K-8*. Thousand Oaks, CA: Corwin Press.

Harlen, W. (2007). *Assessment of learning.* London: Sage.

Harlen, W. (2010). *Principles and big ideas of science education*. College Lane, Hatfield, Herts: Association for Science Education.

Harlen, W. & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1). In: Research Evidence in Education Library. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414-434.

Harris, C. J. & Rooks, D. L. (2010). Managing inquiry-based science: challenges in enacting complex science instruction in elementary and middle school classrooms. *Journal of Science Teacher Education*, 21, 227-240. doi: 10.1007/s10972-009-9172-5

Harrison. C. (2009). Making Assessment Work in Classrooms. In Dillon, J. & Maguire, M. (eds) *Becoming a Teacher*. Milton Keynes: Open University Press, pp. 63-73.

Hautamäki, J., Arinen P., Eronen, S., Hautamaki, A., Kupiainen, S., Lindblom, B., Niemivirta, N., Pakaslahti L., Rantanen, P. & Scheinin, P. (2002). *Assessing learning-to-learn: A framework.* Helsinki: Helsinki University.

Havnes, A., Smith, K., Dysthe, O. & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1) 21-27. doi: 10.1016/j.stueduc.2012.04.001

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

Hickey, D. T., Taasoobshirazi, G. & Cross, D. (2012). Assessment as learning: Enhancing discourse, understanding, and achievement in innovative science curricula. Journal of Research in Science Teaching, 49(10) 1240-1270. doi: 10.1002/tea.21056

Hill, K. T. & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *Elementary School Journal*, 85, 105-126.

Hodgson, C. & Pyle, K. (2010). *A literature review of assessment for learning in science.* Berkshire: The National Foundation for Educational Research in England and Wales. http://www.nfer.ac.uk/nfer/publications/AAS01/AAS01.pdf

Hofer, B. K. (2002). Personal epistemology as a psychological and educational construct: An introduction. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 3-14). Mahwah, NJ: Erlbaum.

Hofstein, A. & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education*, 28, 87-112.

Hoskins, B. & Fredriksson, U. (2008). Learning to Learn: What is it and can it be measured? Luxembourg: Office for Official Publications of the European Communities. doi: 10.2788/83908

Howson, C. & Urbach, P. (1996). *Scientific reasoning. The Baysian approach*. 2nd ed. Chicago: Open Court Publishing Company.

Hur, S. J. (2003). What is scientific literacy? In *A teacher's guide for using web-based resources in the science classroom*. Chapter 1. www.ioncmaste.ca.

Hurd, P. D. (1998). Scientific literacy: New minds for a changing world. *Science Education*, 82(3) 407-416.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking*. London: Routledge Kegan Paul.

Joint Committee on Standards for Educational Evaluation (2003). The student evaluation standards: How to improve evaluations of students. Thousand Oaks, CA: Corwin Press.

Järvelä, S. & Niemivirta, M. (1999). The changes in learning theory and the topicality of the recent research on motivation. *Research Dialogue in Learning and Instruction*, 1, 57-65.

Jenkins, E. W. (1994). Scientific literacy. In T. Husen & T. N. Postlethwait (Eds.), *The international encyclopaedia of education* (pp. 5345-5350). Oxford UK: Pergamon Press.

Jitendra, A. K., Star, J. R., Starosta, K., Leh, J. M., Sood, S., Caskie, G., Hughes, C. L. & Mack, T. R. (2009). Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. Contemporary Educational Psychology, 34(3), 250–264.

Jones, G. A., Langrall, C. W., Thornton, C. A. & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, 32, 101-125.

Jones, G. A, Langrall, C. W., Thornton, C. A & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education*. 30 (5), 487-519. doi: 10.2307/749771

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.

Johnson, R. L., Penny, J. A. & Gordon, B. (2009). *Assessing performance. Designing, scoring and validating performance tasks.* New York: The Guilford Press.

Kishta, M. A. (1979). Proportional and combinatorial reasoning in two cultures. *Journal of Research in Science Teaching*. 16(5). 439-443.

Kind, P. M. (2013). Establishing assessment scales using a novel disciplinary rationale for scientific reasoning. *Journal of Research in Science Teaching*, 50(5), 530-560. doi: 10.1002/tea.21086

Klauer, K. J. (1989). Teaching for analogical transfer as a means of improving problem solving, thinking, and learning. *Instructional Science*, 18, 179-192.

Klauer, K. J. & Phye, G. (1994). *Cognitive training for children. A developmental program of inductive reasoning and problem solving*. Seattle: Hogrefe and Huber.

Klauer, K. J. & Phye, G. D. (2008). Inductive Reasoning: A training approach. *Review of Educational Research*, 78(1), 85-123.

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). The development of national educational standards. An expertise. Berlin: BMBF.

Klopfer, L., E. (1991). Scientific literacy. In A. Lewy (Ed.), *The international encyclopedia of curriculum* (pp. 947-948). Oxford: Pergamon Press.

Koerber, S., Sodian, B., Thoermer, C. & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, 64(3), 141-152.

Kubiszyn, T. & Borich, G. (2010). *Educational testing and measurement. Classroom application and practice* (9th edition). Hoboken, NJ: Wiley.

Kuhn, D., Pease, M. & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology*, 103(3), 268-284. doi:10.1016/j.jecp.2009.01.009

Kuhn, D., Phelps, E. & Walters, J, (1985). Correlational reasoning in an everyday context. *Journal of Applied Developmental Psychology*, 6(1), 85-97.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Kupiainen, S., Hautamäki, J. & Rantanen, P. (2008). EU pre-pilot on Learning to Learn. Report on the compiled data. Helsinki: University of Helsinki Centre for Educational Assessment.

Kush, J. C., Spring, M. B., Barkand, J. (2012). Advances in the assessment of cognitive skills using computer-based measurement. Behavior Research Methods, 44(1) 125-134. doi: 10.3758/s13428-011-0136-2

Laugksch R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84(1), 71-94.

Lawson, A. E., Adi, H. & Karplus, R. (1979). Development of correlational reasoning in secondary schools: Do biology courses make a difference? *American Biology Teacher*, 41(7), 420-425.

Leach, J. T. & Scott, P. H. (2008). Teaching for conceptual understanding: An approach drawing on individual and sociocultural perspectives. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 647-675). New York: Routledge.

Lederman, N. G. & Lederman, J. (2007). Standards for science education in the United States: Necessary Evil? In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it comparable. Standards in science education* (pp. 347-371). Münster: Waxmann.

Leighton, J. P. (2006). Teaching and assessing deductive reasoning skills. *Journal of Experimental Education*, 74(2), 109-136.

Lee, V. S., Greene, D. B., Odom, J., Schechter, E. & Slatta, R. W. (2004). What is inquiry-guided learning? In V. S. Lee (Ed.), *Teaching & learning through inquiry: A guidebook for institutions and instructors*, (pp. 3–16). Virginia Sterling, VA: Stylus.

Levin, H. M. (2013). The utility and need for incorporating noncognitive skills into large-scale educational assessments. In M. von Davier, E. Gonzalez, I. Kirsch & K. Yamamoto (Eds.), *The role of large-scale assessments: Perspective form technology, economy and educational research* (pp. 67-86). New York: Springer.

Limón, M. (2006). The domain generality-specificity of epistemological beliefs: A theoretical problem, a methodological problem or both? *International Journal of Educational Research*, 45, 7-27.

Lissitz, R. W. & Jiao, H. (Eds.) (2012). *Computers and their impact on state assessments. Recent history and predictions for the future.* Charlotte, NC: Information Age Publishing.

Liu, O. L., Lee, H-S & Linn, M. C. (2010). Multifaceted assessment of inquiry-based science learning. *Educational Assessment*, 15(2), 69-86. doi: 10.1080/10627197.2010.491067

Llewellyn, D. (2005). *Teaching high school science through inquiry. A case study approach.* Thousand Oaks, CA: Corwin Press.

Lockwood, E. (2013). A model of students' combinatorial thinking. *Journal of Mathematical Behavior*, 32(2), 251-265. doi: 10.1016/j.jmathb.2013.02.008

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Welsley Publishing Company.

Lott, W. G. (1983). The effect of inquiry teaching and advance organizers upon student outcomes in science education. *Journal of Research in Science Teaching, 20*, 437-451. doi: 10.1002/tea.3660200507

Machamer, P. (1998). Philosophy of science: An overview for educators. *Science and Education*, 7(1), 1-11.

Marzano R. J. & Kendall, J. S. (2007). *The new taxonomy of educational objectives.* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Marzano, R. J. & Haystead, M. W. (2008). *Making standards useful in the classroom.* Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist, 59(1)*, 14–19.

Mayrath, M. C., Clarke-Midurea, J., Robinson, D. H. & Schraw, G. (2012). *Technology-based assessment for 21st century skills. Theoretical and practical implications for modern research.* Charlotte, NC: Information Age Publishing.

McCaslin, M. & Hickey, D. T. (2001). Self-regulated learning and academic achievement: A Vygotskian View. In B. J. Zimmermann & D. H. Shunk (Eds.), Self-regulated learning and academic achievement. Theoretical perspectives (pp. 227-252). Mahwah, NJ: Erlbaum.

MCEETYA (Ministerial Council on Education, Employment, Training an Youth Affairs) (2006). National Assessment Program – Science Literacy Year 6 Technical Report. http://www.mceecdya.edu.au/verve/_resources/NAP_SL_2006_Technical_Report.pdf

McLoughlin, E., Finlayson, O. & van Kampen, P. (2012). Report on mapping the development of key skills and competencies onto skills developed in IBSE. Unpublished manuscript prepared for the SAILS project.

McGuinness, C. (2005). Teaching thinking: Theory and practice. *British Journal of Educational Psychology Monograph, 2*(3), 107-126.

McMillan, J. H. (Ed.). (2013). *SAGE handbook of research on classroom assessment.* London: SAGE.

Misailidou, C. & Williams, J. (2003). Diagnostic assessment of children's proportional reasoning. *The Journal of Mathematical Behavior*, 22(3), 335-368.

Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, 6(2), 91-99.

Molnár, G., Greiff, S. & Csapó (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9*, 35-45. doi: 10.1016/j.tsc.2013.03.002

Muis, K. R. & Franco, G. M. (2009). Epistemic beliefs: Setting the standards for self-regulated learning. Contemporary Educational Psychology, 34, 306-318.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment.

National Research Council/NRC (1996). National science education standards. Washington DC: National Academy Press.

National Research Council/NRC (2000). Inquiry and the national science education standards. Washington DC: National Academy Press.

Nitko, A. J. & Brookhart, S. M. (2011). *Educational assessment of students* (6th edition). Boston, MA: Pearson – Allyn & Bacon.

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

O'Neill, K. & Stansbury, K. (2000). *Developing a standards-based assessment system.* San Francisco, CA: WestEd.

OECD (2003a). *The OECD 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving.* Paris: OECD.

OECD (2003b). *Learners for life. Student approaches to learning. Results from PISA 2000.* Paris: OECD.

OECD (2006). *Assessing scientific, reading and mathematical literacy. A framework for PISA 2006.* Paris: OECD.

OECD (2009a). *PISA 2009 Assessment Framework. Key competencies in reading, mathematics and science*. Paris: OECD.

OECD (2009b). PIAAC problem solving in technology-rich environments. http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=edu/wkp(2009)15

OECD (2012). *Draft PISA 2015 Collaborative Problem Solving assessment framework*. Paris: OECD.

OECD (2013). *PISA 2012 assessment and analytical framework. Mathematics, reading, science, problem solving and financial literacy.* Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52,* 1-12.

Parkes, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 107-123). London: SAGE.

Pelligrino, J., Chudowsky, N. & Glaser, R., (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Piaget, J. (1929). The child's conceptions of the world. New York: Harcourt, Brace and Company.

Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge and Kegan Paul.

Polya, G. (1968). *Induction and analogy in mathematics* (Mathematics and Plausible Reasoning, v. 1). Princeton: Princeton University Press.

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

Pope, M. & Gilbert, J. (1983). Personal experience and the construction of knowledge in science. Science Education, 67(2), 193–203.

Reif, F. (2008). Applying cognitive science to education. Thinking and learning in science and other complex domains. Cambridge, MA: MIT Press.

Renzulli, J. S. & Dai, D. Y. (2001). Abilities, interest, and styles as aptitudes for learning: A person-situation interaction perspective. In R. J. Sternberg & L. Zhang (Eds.), Perspectives on thinking, learning, and cognitive styles (pp. 23-36). Mahwah, NJ: Erlbaum.

Renzulli, J. S. (1977). The Interest-A-Lyzer. Mansfield Center, CT: Creative Learning Press.

Riess, F. (2000). Problems with German scientific education. *Science and Education*, 9(4), 327–331.

Roberts, D. A. (2007). Scientific literacy / Science literacy. In S. K. Abell & N. G. Lederman (Eds.). *Handbook of Research on Science Education.* Mahwah, NJ: Lawrence Erlbaum. 729780.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Henriksson, H. L. & Hemmo, V. (2007). *Science education now. A renewed pedagogy for the future of Europe.* Brussels: Office of Official Publications of the European Communities.

Ross, J. A. & J. Cousins, B. (1993). Enhancing secondary school students' acquisition of correlational reasoning skills. *Research in Science & Technological Education,* 11(2), 191-205.

Roth, K. (1990). Developing meaningful conceptual understanding in science. In B. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction.* Hillsdale, NJ: Erlbaum.

Ruiz-Primo, M. A., Li, M., Wills, K., Giamellaro, M., Ming-Chih, L., Mason, H. & Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. Journal of Research in Science Teaching. 49(6). 691-712. doi: 10.1002/tea.21030

Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3-4), 205-235.

Ruiz-Primo, M. A. & Li, M. (2013). Examining formative feedback in the classroom context. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 215-232). London: SAGE.

Russ, R. S., Coffey, J. E., Hammer, D. & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, 93(5) 875-891. doi: 10.1002/sce.20320

Sato, M., Wei, R. C. & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of national board certification. *American Educational Research Journal*, 45, 669-700.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. Journal of Experimental Psychology, 85, 406-411.

Schroeder, C. M., Schott, T. P., Tolson, H.Huang, T.-Y. & Lee, Y.-H. (2007). A meta-analysis of national research. Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44,* 1436-1460. doi: 10.1002/tea.20212

Schröder, E., Bödeker, K., Edelstein, W. & Teo, T. (2000). *Proportional, combinatorial, and correlational reasoning. A manual including measurement procedures and descriptive analyses.* Study „Individual Development and Social Structure". Data Handbooks Part 4. Berlin: Max Planck Institute for Human Development.

Schwartz, R. & Lederman, N. (2008). What scientists say: scientists' view of nature of science and relation to science context. International Journal of Science Education, 30(6), 727–771.

Schwartz, R. S., Lederman, N. G. & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: an explicit approach to bridging the gap between nature of science and scientific inquiry. Science Education, 88(4), 610–645.

Schwartz, R. S., Lederman, N. G. & Lederman, J. S. (2008). An instrument to assess views of scientific inquiry: the VOSI questionnaire. Paper presented at het annual meeting of the National Association for Research in Science Teaching, March 30-April 2, 2008. Baltimore, MD. http://homepages.wmich.edu/~rschwart/

Shamos, M. H. (1995). *The myth of scientific literacy*. NJ: Rutgers University Press, New Bunswick.

Shayer, M. (1970). How to assess science courses. *Education in Chemistry, 7*, 182-186.

Shayer, M. (1999). Cognitive acceleration through science education II: Its effects and scope. *International Journal of Science Education*, 21(8), 883-902. doi: 10.1080/095006999290345

Shayer, M. & Adey, P. (1989). *Towards a science of science teaching. Cognitive development and curriculum demand.* London: Heinemann Educational Books.

Shayer, M. & Adey, P. (Eds.). (2002). Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 Years. Milton Keynes: Open University Press.

Shymansky, J. A., Hedges, L. W. & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching, 27*, 127-144. doi: 10.1002/tea.3660270205

Simonton, D., K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of Life Reviews*, 7(2), 156-179. doi:10.1016/j.plrev.2010.02.002

Simonton, D. K. (2012). Assessing scientific creativity: Conceptual analyses of assessment complexities. Commissioned paper, The Science of Science and Innovation Policy Conference, National Academy of Sciences.

Sindler, N. W. (2011). *Assessment powered teaching.* Thousand Oaks, CA: Corwin Press.

Schweizer, F., Wüstenberg, S. & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, *24*, 42–52. doi:10.1016/j.lindif.2012.12.011

Stiggins, R. J., Griswold, M. M. & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*. 26(3) 233-246. doi: 10.1111/j.1745-3984.1989.tb00330.x

Suskie, L. (2009). *Assessing student learning. A common sense guide* (2nd edition). San Francisco, CA: Jossey-Bass.

Tamir, P. (1996). Science assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.). *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 93-129). Boston: Kluwer Academic Publishers.

Taylor, A. & Gail, J. (2009). Proportional reasoning ability and concepts of scale: Surface area to volume relationships in science. International Journal of Science Education, 31(9), 1231-1247. doi: 10.1080/09500690802017545

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In Segers, M., Dochy, F. & Cascallar, E. (Eds.*). Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Dordrecht: Kluwer Academic Publisher.

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

Tuan, H-L., Chin, C-C. & Shieh, S-H. (2005). The development of a questionnaire to measure students' motivation towards science learning. International Journal of Science Education, 27, 639-654.

United Nations Educational Scientific and Cultural Organisation [UNESCO] (2001). The training of trainers manual for promoting scientific and technological literacy for all. UNESCO, Bangkok. http://unesdoc.unesco.org/images/0012/001230/123077e.pdf

Venville, G., Adey, P., Larkin, S. & Robertson, A. (2003). Fostering thinking through science in the early years of schooling. International Journal of Science Education, 25(11), 1313-1332.

Venville, G., Rennie, L. J. & Wallace, J. (2009). Disciplinary versus Integrated Curriculum: The challenge for school science. http://www.ias.uwa.edu.au/new-critic

Vosniadou, S. & Ioannides, C. (1998). From conceptual change to science education: a psychological point of view. *International Journal of Science Education*, 20(1D), 1213-1230.

Vosniadou, S. & Kollias, V. (2003). Using collaborative, computer-supported, model building to promote conceptual change in science. In E. De Corte, L. Verschaffel, N. Entwistle & J. van Merrienboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 181–196). Oxford: Elsevier.

Waddington, D., Nentwig, P. & Schanze, S. (Eds). (2007). *Making it comparable. Standards in science education*. Münster: Waxmann.

Watters, J. J. & English, L. D. (1995). Children's application of simultaneous and successive processing in inductive and deductive reasoning problems: Implications for developing scientific reasoning skills. *Journal of Research in Science Teaching*, 32(7). 699-714.

Weinstein, T., Boulanger, F. D. & Walberg, H. J. (1982). Science education effects in high school. A quantitative synthesis. *Journal of Research in Science Teaching, 19,* 511-522. doi: 10.1002/tea.3660190610

Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review,* 11(4), 411-427. doi:10.1023/A:1022065517997.

Wiliam, D. (2013). Feedback and instructional correctives. In J. H. McMillan (Ed.). *SAGE handbook of research on classroom assessment* (pp. 197-214). London: SAGE.

Wise, K. C. & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching, 20,* 419-435. doi: 10.1002/tea.3660200506

Wirth, J. & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice, 10,* 329–345. doi:10.1080/0969594032000148172

Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving – More than reasoning? *Intelligence*, *40*, 1-14. doi:10.1016/j.intell.2011.11.003

Wood, D. (1998). *How children think and learn.* Oxford U.K.: Blackwell.

Wyatt-Smith, C. & Cumming, J. (Eds.) (2009). *Educational assessment in the 21st century. Connecting theory and practice.* New York: Springer.

Zhao, Y. (2012). *World class learners. Educating creative and entrepreneurial students*. Thousand Oaks, CA: Corwin.