1

Developing and Assessing Scientific Reasoning

Philip Adey
Department of Education, Kings' College London

Benő Csapó
Institute of Education, University of Szeged

Introduction

Science education has always been considered to be one of the best tools for cultivating students' minds. Scientific activities such as conducting empirical research, designing and executing experiments, gaining results from observations and building theories are seen as those in need of the most systematic forms of reasoning. The fact that a deep understanding of complex scientific theories requires well-developed reasoning skills leads to the assumption that teaching sciences at school will improve students' thinking skills as well. It probably did in the case of a few students who really deeply understood science, but for the majority this assumption did not work mainly because the science was set too far in advance of students' current cognitive capability so they were unable to engage in it fruitfully.

The argument that learning sciences facilitates the development of thinking was one of the justifications for extending the proportion of science in school curricula. However, the rapid growth of scientific data and their distillation into school curricula often resulted in large quantities of disciplinary content that students were not able to process and understand. Until the second half of the twentieth century, the lack of adequate psychological theories or of evidence-based methods of assessing the effects of science education made it impossible to fulfil the ambitious goals of systematically improving students' reasoning skills.

The gap between the level of abstraction, complexity and organisation of teaching materials on the one hand, and students' actual cognitive development on the other can be narrowed in two ways. One side of the solution is that teaching materials should be better adjusted to students' psychological and developmental characteristics. This requires more information on students' actual developmental level and individualized teaching methods to support students' progress. The other side of the solution is accelerating students' cognitive development in order to elevate their level of reasoning to the requirements of the learning tasks. Research has shown that development can be stimulated by specific activities and exercises, and learning science offers a number of efficient opportunities to accelerate students' cognitive development (Adey & Shayer, 1994). Systematic monitoring of the development of students' reasoning skills may facilitate both directions of this adjustment (Glynn, Yeany & Britton, 1991).

In this chapter, first we summarise the results of psychological and educational research concerning cognitive development related to science education. Next, we systematically describe what thinking processes might be developed in science education. Then we illustrate the possibilities by introducing some of those methods which utilise these results in science education and aim at more efficient training of students' thinking processes and finally discuss how these thinking processes can best be measured, diagnosed and monitored in order to support teaching and learning.

Reasoning in Science: Cognitive Development in an Educational Context

Science Reasoning and General Reasoning

Is scientific thinking special? That is, is scientific thinking distinctly different from thinking in other subject areas? Obviously, there are some special characteristics, but to what extent are these simply particular expressions of the human ability to process information in general? Human cognition and the accumulation of experiences are often comparred to the process of scientific research and discovery. However, although there are broad analogies between the logic of scientific research and human reasoning, there are some significant differences as well (Howson & Urbach, 1996; Johnson-Laird, 2006). One of the major differences stems from the developmental nature of human cognition. Humans reach their actual reasoning capacity through a long developmental process, which is shaped by the stimuli and information one has received and processed. Although science has also reached its current form through a long developmental process, the logical system that children are expected to comprehend is a stable constant structure, while children attempting to master it may be in different developmental stages.

Certainly Jean Piaget and his co-workers regarded scientific thinking as representative of general intellectual processing, or general intelligence. During investigations of children's development of thinking from infancy to adolescence, they used practical tasks such as ordering things by size, exploring conservation, cause and effect, control of variables and probability (e.g., Inhelder & Piaget, 1958; Piaget & Inhelder, 1974, 1976), all of which would be easily recognised by mathematics and science teachers as central to their subject areas. He drew conclusions about cognitive development in general from children's performance in these apparently scientific tasks. Also, typical non-verbal tests on general intelligence such as Raven's Matrices (Raven, 1960) or the Calvert Non-verbal test (Calvert, 1986) tap into subjects' ability to use inductive and deductive reasoning which is the basis of a much scientific thinking.

On the whole, this extrapolation from scientific thinking to thinking in general has received some empirical support. Although the general stages of cognitive development described by Piaget are expressed in scientific terms, their descriptions in terms of concrete operations or abstract reasoning are easily applied across all forms of learning. Furthermore, as we will describe later in this chapter, training in scientific thinking has been shown to transfer to higher levels of achievement in remote subject areas such as native or second language learning (Csapó & Nikolov, 2009) suggesting, at least, an intimate link between science reasoning and reasoning in general.

Notwithstanding such evidence it is possible to make some distinction between scientific thinking and 'good' thinking in general. Consider this list of general thinking skills (from McGuinness, 2005):

(1) pattern-making through analysing wholes/parts and similarities/differences;
(2) making predictions and justifying conclusions;
(3) reasoning about cause and effect;
(4) generating ideas and possibilities;
(5) seeing multiple perspectives;
(6) solving problems and evaluating solutions;
(7) weighing up pros and cons;
(8) making decisions.

The first three have ready expressions within science. The fourth, that is, generating ideas, is certainly important in science, but – in a different guise – it is also central to artistic and literary creation. The fifth – seeing multiple perspectives – may be necessary at the frontiers of science for trying to integrate apparently conflicting models (e.g., wave-particle duality). However, at school level it is not as typical of science as it would be of, say, history, social studies or drama where high level thinking includes the ability to see events from a number of different perspectives. It may also be imbued with an emotional load (*can* I see the viewpoint of my enemy?) which is, at least theoretically, less common in scientific thinking. Notwithstanding, it may be important in teaching: teachers should often try to observe a phenomenon from a child's point of view in order to understand the way children reason and that they draw conclusions differently in comparison with an expert. The last three are certainly very general and apply far beyond the boundaries of the sciences. In particular 'solving problems' is something of a catch-all phrase which can embrace many activities. When, as within PISA frameworks, the idea of complex problem solving is well-characterised (OECD, 2003), it is seen as much broader than a scientific ability.

On this argument science education seems to have less to offer in the development of general reasoning ability. Yet, our final conclusion on the debate about the generality-specificity of thinking must rest on the model of intelligence that is adopted. If each of the thinking skills is relatively independent of one another, then each needs to be developed in its own right. On the basis of this model, it is possible to conceive of an individual who scores high on reasoning about cause and effect but low on decision-making. The alternative is to regard each of the individual thinking skills as expressions of a general underlying intelligence. In this case, work on developing a sub-set of whichever list of thinking skills we happen to favour should have some transfer effects to those skills not explicitly trained.

Elsewhere, (Adey, Csapó, Demetriou, Hautamäki, & Shayer, 2007) we have argued that there is indeed a general intelligence, which is amenable to educational influence offering a potential mechanism by which thinking abilities may be transferred from those trained to others. This model also posits that 'on top' of this general processor (g) there exist a set of specialised structural systems (Demetriou, 1993) which allow for a limited independent variation of different areas of thinking (e.g., quantitative-relational, spatial). A critical feature of this model is that the development of the specialised systems is both limited by and is the route into the development of the general intellectual processor and its executive control (self-regulation). We believe that there is substantial empirical evidence which is compatible with this model and that it offers a fruitful basis for educational action and for the analysis offered in this chapter.

Learning and Development

Discussing the problem of development in educational context it is necessary to clarify its relationship to learning. The distinction between 'learning' and 'development' is one about which Vygotsky was exercised at some length. Vygotsky thinks that formal education in one specific domain definitely influences development in other domains of knowledge by a sort of generalisation process… (Tryphon & Vonèche, 1996. p. 6). Indeed, the whole idea of the *Zone of Proximal Development* can be seen as Vygotsky's attempt to explain the relationship between learning and development.

Although we cannot make a sharp distinction between the two concepts, it may be possible to characterise extreme (stereotypical) examples of each term. At the limits, one thinks of 'learning' in relation to content matter and the acquisition of simple knowledge such as the correct spellings of words or multiplication tables, whilst 'development' relates to functions which unfold during a process of maturation, minimally or not at all influenced by the environment. Development is an organic process; a certain stage is based on the previous ones.

Of course, in reality there can be no such thing as 'pure' examples of learning or development in these stereotypical terms – learning uninfluenced by development, or development uninfluenced by experience. Erroneous belief in such stereotypes is at the root of much misunderstanding in education, for example, cognitive development or the unfolding of intelligence is entirely under the control of time and heredity, or that the acquisition of concepts requires only sufficient effort of learning regardless of their inherent complexity.

This problem may be illustrated by an example taken from mathematics education. Hungarian students learn how to convert hours into minutes, meters into millimetres etc. by the fourth grade with considerable effort of memorising the rules and mechanically exercising the conversion operations. Then, they pass to the next chapters of curriculum, learning of conversion ends, and they begin to forget what they have learnt. Their proportional reasoning is at a lower developmental level at that age, and learning rules of conversion has a little impact on it. Later, on the other hand, by the seventh grade they can convert measures again quite well, as it is a specific application of proportional reasoning that reaches a higher developmental level by that time (Csapó, 2003).

Several empirical studies demonstrated that learning sciences does not result necessarily in better scientific reasoning. For example, Bao et al. compared Chinese and American university students' physics knowledge and scientific reasoning. They have found that although Chinese students performed much better on the science knowledge test (attributable to their more demanding high school science studies), their performance on the science reasoning test was similar to that of their American peers (Bao et al., 2009).

It is more useful to see learning and development as lying at either ends of a spectrum, with the simple acquisition of knowledge at the L-end (but still dependent to some extent on the individual's level of maturity) and the development of general intelligence at the D-end (but still amenable to educational stimulus). The acquisition of complex concepts (e.g., photosynthesis or multiple causes of historical events) lies part way along the L-D-spectrum since they develop in complexity in an individual over many years while being strongly under the influence of learning experiences. As far as this chapter is concerned, the development of scientific reasoning is another example of a process which depends on both the development of the central nervous system (the individual's capacity to process complex ideas) and appropriate learning experiences. High-level learning cannot take place without development, and satisfactory cognitive development cannot occur without appropriate cognitive stimulation (learning experiences).

A feature of this Learning-Development-spectrum worth noting is that the *generality* of functions increase as one moves from L to D. At the L-end information learnt tends to be specific and applicable to a narrow range of cognitive functioning. Learning the number of a bus for a particular route is not knowledge that generalises usefully to other contexts. On the other hand, educational experiences which stimulate the development of general intelligence may be expected to have an impact on the effectiveness of all learning, in any intellectual field (and maybe beyond).

The model of a plastic general intelligence proposed here, that is, a general thinking machinery amenable to educational influence, has implications for the whole nature of education. We will return to the question of how science educators can use this model to provide general cognitive stimulation for their students, but now we must consider in more detail some different types of thinking in science which might form the 'subject matter' of a strand in the curriculum devoted to the development of scientific – and by the way, general – thinking.

A System of Thinking Processes That Should Be Developed in Science Education

The processes of thinking have been studied, described and categorised in several psychological and educational research traditions. These approaches often used different theoretical frameworks, terminologies and methods. Among these is the psychometric approach (intelligence research, individual differences approaches, factor analytic studies) which produced a great amount of data of the general cognitive abilities and also contributed significantly to the development of psychological testing and educational assessment (Carroll, 1993).

Piaget and his colleagues emphasised the developmental aspects of cognition, and described the development of thinking through qualitatively different stages. Piaget's work is especially important for science education as his theory explains the origin of reasoning schemes and makes a connection between the manipulation of external objects and the development of higher-order thinking skills. His work has been followed by several Neo-Piagetian researches proposing a number of elaborated models of cognitive development and systems of thinking (e.g., Demetriou, 2004). Piaget's theory and the researches of his followers are especially important for establishing early science education, organising observations and experiments to be carried out by children.

The information processing approach emphasised the differences between novices and experts in the organisation of knowledge. It offers useful models of learning within the content domains, but developmental aspects and reasoning processes are less elaborated in the information processing paradigm. The most recent cognitive neuroscience research studies thinking from another aspect. Its results are not ready for direct application in the field of science education, but the main messages of the results for education are promising: they confirm the claim of the plasticity of the brain and the modifiability of cognitive processes, especially during the early phases of the development (Adey, Csapó, Demetriou, Hautamäki, & Shayer, 2007).

For assessing scientific reasoning we may provide a framework from all these research traditions. However, taking the developmental aspects, the target age groups and the diagnostic orientation into account the Piagetian tradition offers the most useful resources.

There are very many ways in which the cake that we call 'thinking' may be sliced up. In the next section we will first look at a couple of meta-strategies for thinking about thinking, then consider a number of quite general classes of thinking, and then of *dichotomies*. Finally, we will focus on a specific set of 'reasoning patterns' which have particular relevance to science.

Meta-Strategies and General Thinking Processes

Human thinking, in broader practice is never a simple mechanical process. It is always influenced by the actual situation and context as well as the general psychological state of the thinker. Even scientific thinking is often mediated at least at the level of general thinking processes by non-cognitive factors such as motivation, interest and curiosity. Forming science-related attitudes and values may be an important goal of science education, as is the development of beliefs related to the validity of scientific knowledge and the way students think about the status of their own knowledge (personal epistemologies). We will not deal with the affective aspects of learning science in detail in this chapter, but here at the outset we have to mention the possible connection between cognitive and affective processes.

Meta-strategies relate to a person's control over their own thinking process. To some extent they are dispositional but they regulate the whole process of thinking including attention and the choice of deployment of one or another specific types of thinking. There are several research directions which deal with these questions. Meta-cognition is the broadest concept; beyond its importance in scientific reasoning it plays an important role in reading comprehension and mathematical problem solving as well (Csíkos, 2007). These meta-strategies are essential in learning sciences, especially in understanding and mastering complex scientific concepts and ideas.

There are some general thinking processes that are characteristic of some contexts and situations, such as argumentation and critical thinking. It is worth briefly defining them here as well.

Storage and Retrieval

Knowledge about the processes of remembering, also called meta-memory, is more specific than the general processes of self-regulation. These are skills that can be learnt enhancing the thinker's ability to transfer information to and from long-term memory. As human memory stores organised information more efficiently than independent pieces of information, information should be arranged into compact structures before memorising. If the knowledge has a natural structure the best way is to make this structure explicit and the related pieces of information should be memorised by integrating them into this structure. If a unifying structure does not exist, the learner has to create an artificial one and integrate the information into it. For example, a well-known strategy is associating a list of words to be memorised with the parts of a popular building or the houses of a familiar street (method of places). Students with good memorising abilities are able to distinguish between well-structured learning materials when exploring and understanding may result in meaningful conceptual learning, from unstructured information where creating artificial structures may be a better strategy. Storage and retrieval strategies were already studied by Greek philosophers and special techniques (also referred to as *mnemotechnics*) were further developed by the Roman orators.

Self Regulation

This means the ability to attend to the relevant parts of a problem, to analyse personal reasoning and monitor one's own choice of thinking pathways, progress towards a solution and detection of errors and dead-ends. Self regulation includes motivational and other affective aspects as well (Molnár, 2002).

Argumentation (Dialogic)

Dialogic argumentation identifies disagreement among assertions, relates supporting and refuting evidence to each assertion, and weighs all of the evidence "in an integrative evaluation of the relative merit of the opposing views" (Kuhn, 1992, p. 157). Argumentation plays a relevant role in the advancement of science by checking errors and identifying insufficient evidence. Argumentation requires organising statements into a logical order. It is a basic reasoning process in presenting the results of a research, but its potential is not yet fully exploited in science education (Osborne, 2010).

Critical Thinking

Critical thinking belongs to those forms of thinking which are most frequently mentioned both inside and outside the school context. Its improvement is frequently proposed, recently due to the explosion of easily accessible information. One often has to select and classify information and has to evaluate its relevance and validity and has to judge the credibility of its sources. At the same time, definitions of critical thinking are generally difficult to operationalise. The core of critical thinking is usually identified as the ability of collecting, organising and evaluating information.

Most interpretations describe critical thinking as a set of a number of component abilities, and the long lists of components usually include every important form of thinking. The most frequently mentioned attributes of critical thinkers are openness, the intention of checking the reliability of information sources, assessing the foundation and validity of conclusions, evaluating the quality of arguments and the ability of questioning (Norris & Ennis, 1989; Ennis, 1995).

If we look for the distinctiveness of critical thinking, the feature that makes it more than the sum of its components, we find it the way the process of thinking is organised and in its purpose. There is always a strong critical attitude behind a critical thinking act that motivates the thinker to question a given bit of information, statement, model, theory, chain of arguments etc. Thinking processes mobilised by critical attitudes play an essential role in the advancement of science, especially in evaluating results, judging evidence, filtering out sources of errors, and falsifying unjustified statements. Preparing critical analyses and reviews is one of the characteristic activities of the researcher. Science education offers an efficient field for practising critical thinking as the validity of arguments may be judged on the basis of objective criteria.

Dichotomies

Some forms of thinking relevant to science may be characterised by dichotomies, introduced briefly in this section. In few of the following pairs there is not any question of one being 'better' than the other. In all except the case of concrete-abstract, the highest level of thinking involves an integration of both types, or a choice of the most appropriate type for a particular situation.

Quantitative – Qualitative

Quantitative reasoning is characterised by situations where the learner must apply properties and procedures related to number sense and number operations to solve the given problem. Qualitative thinking focuses more on the nature of the variables and judgement for the purpose of comparison or prioritising. In most complex problem-solving situations both quantitative and qualitative reasoning need to be employed.

Concrete – Abstract

Concrete thinking is restricted to actual objects, words, or numbers and simple relationships between them. It allows for simple mathematical manipulation, classification and simple causal relationships. Abstract thinking allows for the imaginary manipulation of factors in a hypothetical model or the possibility of understanding complex relationships such as when there are multiple interacting causes and multiple interacting effects. In this case, there is a clear hierarchy with abstract thinking being far more powerful than concrete thinking. As from abstract constructs further abstract ones can be created, understanding complex systems may require the comprehension of several levels of abstraction. Science offers an excellent context for developing abstraction skills and for demonstrating the concrete-abstract relationship and levels of abstraction.

Convergent – Divergent

Convergent reasoning is used in the type of problem which has one correct answer, so that the reasoning progresses through steps designed to reach this one answer. These steps may include the elimination of extraneous variables, the combination of others, and operations on given data with the aim of reaching the correct solution. Divergent thinking by contrast is discursive, exploring a number of solutions, especially to problems which may have more correct answers. Divergent thinking is also characteristic of creativity, 'thinking outside the box' and 'lateral thinking'. Complex problems may require both divergent and convergent thinking in different phases of their solution.

Wholist – Analyst

The wholist-analytic dichotomy represents a general approach to a problem or to representing and processing information, also identified as cognitive style (Davies & Graff, 2006). Wholist thinking aims for an overview of a situation, to reach a conclusion based on the 'big picture' rather than the detail. The opposite, analytic approach is to focus on the detail and try to solve the problem bit by bit. Analytic thinking is characterised by situations where the learner must apply principles from formal logic in determining necessary and sufficient conditions or in determining if implication of causality occurs among the constraints and conditions provided in the problem stimulus. Excessive wholist thinking may miss important details, and excessive analytic thinking may fail to integrate the parts of a solution into a coherent response. Both types of thinking are useful at appropriate phases of problem-solving. (Note that some authors use 'holist' rather than wholist.)

Deductive – Inductive

The process of deduction is reasoning from the general to the specific or from premises to a logically valid conclusion. Examples are: Conditional (deducing a valid conclusion from a rule of the form "if P, then Q"); Syllogistic (evaluating whether a conclusion necessarily follows from two premises that are assumed to be true) or more generally Propositional reasoning; and Suppositional (Supposing a possibility for the sake of argument, in some cases obtaining a contradiction). Deductive reasoning applies strict logical rules. Consequently, appropriate application of rules to true premises always results in true conclusions. On the other hand, deductive reasoning does not produce originally new knowledge as it expresses in a different form what is there already, although often in a hidden way in the premises. Deductive reasoning is essential in scientific research, errors in a deductive process leading to false conclusions. As Piaget's research demonstrated, children attain a fully developed formal logical system only after a long developmental process (and we may add: if at all), therefore they possess limited tools to comprehend deductive argumentation. (For the development of deductive reasoning and its relevance for science education, see Vidákovich, 1998).

The process of induction is reasoning from particular facts or individual cases to a general conclusion, that is, constructing a general rule or explanatory model from a number of specific instances. Classically, science progresses by a series of inductive and deductive loops, although this rather convergent picture omits the intuitive, creative leap that very often occurs in real scientific advance. From a philosophical point of view, accumulation of positive examples may not prove the truth of a theory in general, therefore, Popper proposed a more sophisticated theory for explaining induction that is based on the concept of falsification (Popper, 1972). Psychological processes of inductive reasoning play significant role in understanding science and application of knowledge in new contexts (Csapó, 1997, 2001a). Its modifiability has been demonstrated in a number of training experiments (Hamers, de Koning, & Sijtsma, 1998; Sanz de Acedo Lizarraga, Sanz de Acedo Baquedano, & Oliver, 2010. Molnár, 2011).

Thinking Patterns, Operations, Abilities

Finally, in this section on taxonomies of thinking we will look at a number of specific reasoning patterns, or 'schemata' which appear to be characteristic of scientific thinking. A variety of terms have been used as comprehensive names for them; for example, patterns, schemes, schemata, operations, skills and abilities. We acknowledge that several terms may be appropriate in different contexts; however, we prefer to use *thinking abilities* as the most general term for them. We note again, that we consider them as plastic abilities, modifiable by systematic educational stimulation.

They vary in the demand they make on intellectual capacity and here they are ordered very approximately in terms of their difficulty. Because these abilities are really aspects of general cognitive development, they are not amenable to direct instruction, but need to be slowly constructed by students in response to maturation and appropriate stimulating experiences.

Piaget and his colleagues studied the development of these reasoning operations by observing children's activities dealing with simple tasks related to scientific phenomena (see Inhelder, & Piaget, 1958; Piaget & Inhelder, 1974, 1976). Other researchers studied them by the means of mental tests. The development of some of these operations was assessed in several projects in Hungary by paper-and-pencil tests (see Csapó, 2003).

Conservation

For an adult it is obvious that a quantity (of matter, number etc.) remains the same if nothing is added or taken away from it. Conservation is the result of development appearing at a certain stage. Before it a child does not recognise that changing insignificant features, e.g., the pouring water from one cup into another one with a different shape does not influence the quantity of the water. Conservation of number (two rows of beads are still the same number when one is stretched) is one of the simplest forms of conservation while recognising that a solid displaces an equal volume of liquid in which it sinks is more demanding.

Seriation

This means not only putting things in order according to one or more properties, but also interpreting a given phenomenon within a series of comparable phenomena in order to assign some plausible meaning to it. E.g., ordering stimuli along a quantitative dimension, such as length (Inhelder & Piaget, 1958; Nagy, 1987). Seriation is a precondition for solving more complicated organising tasks, e.g., trying all setting of an experiment.

Seriation, in general dealing with relations is an essential feature of scientific reasoning. *Transitivity* is a feature of relations frequently necessary to handle. In general, transitivity involves the ability to understand the characteristics of relationships and logically combine two or more relations to draw a conclusion. Combining two or more relations leads to identifying new or more general relations (Glenda, 1996).

Classification

Classification is the ability to classify objects or ideas as belonging to a group and having the characteristics of that group. At its simplest, this may demand no more than grouping objects which have just one variable with two values. ("Group these red and blue squares so that all in each group are the same."). As the number of variables and values increases so does their difficulty, and extra layers of demand are added by empty classes, class inclusion (two classes in which all members of one class are included in the other, as in the proposition "All dogs are animals") and two-way classification. ("Lions are mammals within vertebrates within animals but they are also carnivores.") More complex structures require multiple classification and hierarchical classification (Inhelder & Piaget, 1958; Nagy, 1987).

Combinatorial Reasoning

Combinatorial reasoning is the process of creating complex constructs out of a set of given elements that satisfy the conditions explicitly given or inferred from the situation. This is characterised by situations where the learner must examine a variety of factors, consider all combinations in which they can appear, evaluate each of these individual combinations relative to some objective constraint and then select from or rank the combinations into order. If the conditions and constraints allow a larger number of constructs, all constructs can be created only if a systematic order of enumeration is applied (for a taxonomy of combinatorial operations, see Csapó, 1988; for developmental data see Csapó, 2001b; Nagy, 2004). Creating combinations of conditions or values of variables systematically is often required when designing experiments (Inhelder & Piaget, 1958; Kishta, 1979; Schröder, Bödeker, Edelstein, & Teo, 2000). Physical and chemical experiments offer a great number of possibilities to exercise combinatorial reasoning by exploring all possible settings allowed by the constraints of the equipment and materials. (For the improvement of combinatorial reasoning see also Csapó, 2003.)

Analogical Reasoning

Analogical reasoning can be applied in situations where the learner must solve a problem with a context similar to a problem the learner is familiar with or includes a problem base which the learner has solved in the past. The parameters or the context in the new stimulus material is changed, but the driving factors or causal mechanism is the same or similar. The learner should be able to solve the new problem by interpreting it in the light of past experience with the analogous situation. Where the reality and the analogy are both accessible to direct perception, we refer to this as concrete modelling (for example the notion of temperature rising is modelled by the thread of mercury rising in a thermometer) but where either or both are abstraction, it becomes formal modelling (relating potential difference to water pressure). Analogical reasoning relates two individual objects or phenomena based on their structural similarities. Analogical reasoning is one of the basic mechanisms of transfer and the application of knowledge (Klauer, 1989a). Finding similarities between more than two objects, and analysing the rules of similarities lead to rule induction and inductive reasoning (Polya, 1968). Analogical reasoning helps understanding new scientific phenomena on the basis of already known similar phenomena, as well as application of knowledge in new areas. Therefore, learning science offers several possibilities of improving analogical reasoning (Nagy, 2006).

Proportional Reasoning

Proportional reasoning involves a sense of co-variation and of multiple comparisons, and the ability to mentally store and process several pieces of information. The co-variation is usually assumed to be linear, but in general could be non-linear (e.g., exponential); considering a nonlinear as a linear relationship may lead to oversimplification or a serious thinking error. Proportionality requires the comparison of two or more ratios (Schröder, Bödeker, Edelstein, & Teo, 2000). Proportional reasoning is a basic process involved in several more complex analogical and inductive forms of reasoning (Csapó, 1997). Understanding some basic scientific concepts (e.g., speed) requires proportional reasoning, and one of the obstacles of understanding school science is the lack of a proper level of proportional reasoning (Kishta, 1979). Recent research has also demonstrated that although proportional reasoning develops over a long period (Boyera, Levinea, & Huttenlochera, 2008), it is amenable to training (Jitendra et al., 2009).

Extrapolation

Extrapolation enables learners to use the pattern of data from one area to predict what will happen in another area. Extrapolation is closely related to analogical and inductive reasoning while rules induced from observation in one area are applied to another area not directly explored. In simple cases, extrapolation means extending the scope of relationships beyond the range of measured data or creating new data points. In more general cases extrapolation requires extending complex rules to new, unknown situations. The probability of making errors and invalid extrapolation increases with the distance between the observed and extrapolated data or rules.

Probabilistic Reasoning

Most scientific phenomena as well as events of everyday life depend on probability. There is always a certain probability that it is raining in a given day; that a team wins a given match; or that the exchange rate of a given currency will change. Understanding these phenomena and calculating risks require probabilistic reasoning. Probabilistic inferences are based on past events and assumed (or calculated) likelihoods of future events. Risk analysis depends on this, and the realisation that one or several counter examples do not undermine the validity of an established probabilistic relationship. Development of probabilistic reasoning was studied by Piaget mostly in the context of simple science experiments (Piaget & Inhelder, 1975; Girotto & Gonzalez, 2008).

Correlational Reasoning

Correlational reasoning means dealing with probabilistic relationships when the connection between two features or variables appears only in certain number of cases. Depending on the ratio of the appearances, the strength of the association may be different. Recognising correlational relationships involves observation of cases confirming and not confirming the association, and estimating their ratio (Kuhn, Phelps, & Walters, 1985; Schröder, Bödeker, Edelstein, & Teo, 2000). As it requires observations, collecting and processing contradicting information, mastering correlational reasoning is seldom complete, and its failures may lead to doubtful judgements (Bán, 1998). Research has shown that it develops slowly (Lawson, 1982; Koerber, Sodian, Thoermer, & Nett, 2005), but it can be improved with systematic instruction, especially in science (Lawson, Adi, & Karplus, 1979; Ross & Cousins, 1993).

Separation and Control of Variables

Control of variables is a complex reasoning pattern or strategy which may involve several other simpler reasoning schemes. It is a result of a long developmental process and is reached during the formal reasoning phase. During an early developmental phase, children learn to identify the key components of a system (e.g., the string and the ball in a pendulum), associate variables with them (e.g., length and weight), and differentiate between the values of the variables (e.g., short, long; light, heavy). Investigating the connection between the variables, and determining their dependencies requires systematic manipulation of the variables, changing their values and observing their effects on the others. Control of variables is essential in designing scientific experiments, organising and interpreting results of observations.

Advancing Cognitive Development through Science Education

In the last section we described in some detail a set of thinking abilities which are important in science – but in the first section we intimated that scientific thinking is rooted in general thinking ability, and that the development of one is likely to transfer to the other. Now we must address the question of by what mechanism can students' scientific reasoning (and by extension all of their reasoning) be stimulated? We have made it clear that we do not subscribe to a 'fixed intelligence' viewpoint, but believe in (and have good evidence for) a model of general and specific thinking that is amenable to educational influence. On the Learning-Development spectrum introduced in a previous section, reasoning falls nearer to the Development-end. In other words it is more developmental, and more general than a simple learning task and we should not expect that scientific reasoning (for example the thinking abilities described in the last section) could be taught in a direct instructional manner. Any attempt to 'teach' them as a set of rules to be followed is doomed to failure. The student may memorise the rules but fail to internalise them, to make them his/her own, and it will mean that s/he will be lost when trying to apply the rules. The development of scientific reasoning, as with the development of any reasoning, must necessarily be a slow and organic process in which the students construct the reasoning for themselves.

We now need to say more about what the teacher can do to facilitate this process of construction. We will exemplify the general principles with reference to one particular approach, that of Cognitive Acceleration through Science Education (CASE), and then conclude this section by mentioning briefly how similar principles are employed by a number of other successful programmes for the teaching of thinking. CASE is chosen as the prime exemplar since it has been well-established over a period of 20 years originating from a science context, and has published many examples demonstrating the effectiveness of its approach (Adey, Robertson, & Venville, 2002; Adey & Shayer, 1993, 1994; Shayer, 1999; Shayer & Adey, 2002).

CASE pedagogy is founded in the developmental psychologies of Jean Piaget (1896-1980) and Lev Vygotsky (1896-1934). Whilst they had arguments over some important issues during their lifetime (such as the primacy of language over development or development over language), they agreed about many things, notably:

(1) the impact of the environment on cognitive development;

(2) the at least equal importance of the social as well as the physical environment;

(3) the value to children's development of becoming conscious of their own thinking processes, conscious of themselves as thinkers.

These three principles are the basis of what are called the 'pillars' of cognitive acceleration. Firstly, the specific nature of a stimulating environment is one that is challenging, one that goes beyond what an individual is currently capable of, one that requires intellectual effort to tackle. In Piagetian terms this would be called *Cognitive Conflict*, and for Vygotsky it is working within the *Zone of Proximal Development* – the difference between what a child can do unaided and what they can achieve with the support of a teacher or more able peer. According to Vygotsky, the only good learning is that which is in advance of development (Vygotsky, 1978). The task for the teacher, which is not trivial, is to maintain just the right degree of tension between what her students can manage easily and what they will be incapable of at this stage, no matter what support they receive. This task is made even more difficult when, as is usual, a class contain students of a wide range of cognitive levels. An activity which offers cognitive conflict for one student may seem trivial to another, and impossibly difficult to a third. Activities which are generative of cognitive stimulation for classroom use must have a variety of entry points and an increasing slope of difficulty so that all can make a start, and all encounter some challenge along the way.

Secondly, lessons which promote scientific reasoning provide plenty of opportunities for *social construction*. That is, they encourage students to talk meaningfully to one another, to propose ideas, to justify them, and to challenge others in a reasonable manner. A stimulating classroom is characterised by high-quality dialogue, modelled and orchestrated by the teacher. Those students who are just a few steps ahead of their peers may be especially efficient helping the others as they think in similar way and are sensitive to the obstacles of understanding.

Thirdly, classrooms in which reasoning is being developed are reflective places. Students and the teacher look back on the thinking they have developed and reflect on successes and failures, so that the lessons of the development of a particular reasoning strand can be learnt and transferred to future 'thinking' lessons. Metacognition encourages the abstraction of general reasoning principles which can subsequently be applied to new types of reasoning.

In cognitive acceleration these three core 'pillars' were originally incorporated into a set of 30 activities aimed at junior secondary students aged 11–14 years (Adey, Shayer, & Yates, 2001) but the principles have now been applied to a younger range of children (Adey, 1998; Adey, Nagy, Robertson, Serret, & Wadsworth, 2003; Adey, Robertson, & Venville, 2001). In all cases, schemata of reasoning such as those described in the last section form the 'subject matter' of the activities. For example, starting with the schema of *classification*, in one activity students aged about 7 years are presented in their groups with a collection of seed-like objects including an apple pip, sunflower seeds, a rice grain, small glass beads, lentils, raisins and so on. They are asked to study them and say which are seeds and which are not. Making piles of seeds and not-seeds is easy enough but now they are asked to justify their choices. This leads to much discussion, carefully led in an open-ended manner by the teacher, generating cognitive conflict as the class struggles together towards some set of features by which a seed can be distinguished from a non-seed.

With the youngest children such activities are given about 30 minutes every week, while with the 7 to 9 year olds perhaps activities last an hour and are given once every two weeks over two years. Evaluations (Adey et al., 2002; Shayer & Adey, 2002; Shayer & Adhami, 2011; Venville, Adey, Larkin, & Robertson, 2003) show that such intervention has long term effects on the development of children's reasoning which transfers to gains in achievement in academic subject areas.

Other programmes which have reported significant effects on children's reasoning include *Philosophy for Children* (Lipman, Sharp, & Oscanyan, 1980; Topping & Trickey, 2007a, 2007b). Although this training does not have a particular focus on science, the classroom methods applied in this program (interaction between students, discussion, argumentation) may be useful in science education as well. Similarly, science-related philosophical questions may be discussed in this way; furthermore students' attitudes, beliefs and personal epistemologies may be efficiently formed by this approach. (For the Hungarian adaptation of the *Philosophy for Children* program, see G. Havas, Demeter, & Falus, 1998.)

Another training method for fostering thinking relevant to the education of sciences is Klauer's Inductive Reasoning Program (Klauer, 1989, 1996; Klauer & Phye, 1994, 2008). Originally, the program applied a toolkit designed on the basis of Klauer's model of inductive reasoning (Klauer, 1998b). It proved to be especially effective with young slow-developing students. Later these principles of development were applied both outside the particular school subjects and embedded into them. In a recent experiment, based on Klauer's original model, Molnár (2011) reported successful fostering of inductive reasoning in young children by using playful but well-structured activities. In a current article, Klauer and Phye (2008) reviewed 74 developmental studies which aimed at improving inductive reasoning. Most of the interventions took place in the framework of school subjects, including mathematics, biology, geography, and physics.

Several further experiments demonstrated that science education offers excellent opportunities for fostering thinking abilities. Among others, Csapó (1992, 2003) reported significant improvements in combinatorial reasoning as a result of training embedded in physics and chemistry. Nagy (2006) described an experiment aiming at fostering analogical reasoning in biology that not only improved analogical reasoning but resulted in better understanding and mastery of biology content as well. Beyond the experimental works and intervention studies, this approach – embedding developmental effects into the delivery of science content – may be applied in regular everyday teaching as well. For example, Zátonyi (2001) proposes a number of particular activities for physics education which may serve multiple aims, fostering thinking abilities and a better mastering of the content.

There are several teaching methods which are especially favourable for the advancement of thinking. A recent movement promoting *Inquiry Based Science Education*[1] (IBSE) proposes more observations and experiments in science education. *Problem Based Learning* (PBL) organises teaching materials around realistic issues, often cutting across disciplinary borders, which indicate the relevance of learning specific pieces of information. Dealing with complex problems is not only more chal-

1 IBSE is the model that is supported by European Federation of National Academies of Sciences and Humanities and its Working Group Science Education, see: http://www.allea.org/Pages/ALL/19/243.bGFuZz1FTkc.html. A number of European Commission projects deals with IBSE as well.

lenging but more motivating for young learners as well, compared to the often sterile materials organised by the disciplinary logic. Project work also requires more activities fostering thinking, and helps to integrate knowledge into context. Group projects especially foster communication skills and group problem solving.

Assessing Cognitive Development in Science Education

Assessing reasoning requires tools and methods different from that of assessing how well students learnt content knowledge. The main problem is that assessing thinking always requires content and the familiarity of content may influence the related reasoning and the solution itself. Piaget faced a similar difficulty when he studied children's reasoning processes. Therefore, he applied a method of questioning the students – the *clinical method* – which provided most of the information needed by the examined child so focusing the test on the ability to use and process information. A similar problem has to be solved when assessing thinking: the influence of the content should be minimised.

Content of Assessment

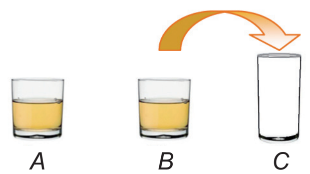
When we are assessing science reasoning, we are by definition not assessing science knowledge, even science conceptual knowledge. The task therefore becomes one of trying to measure a student's ability to reason scientifically while making the least possible demands on their content knowledge. If an item confounds knowledge and reasoning and a student fails, we do not know whether that failure represents a lack of knowledge or inadequate reasoning powers. While it is probably impossible for a reasoning item to demand no knowledge at all (or indeed for a knowledge item to require no reasoning at all), that at least is the ideal to strive for. What knowledge is needed should be provided. For example, if we wish to assess a young child's ability to conserve liquid volume across change of shape we might present an item such as the one on Figure 1.1.

Here are two glasses A and B. They are just the same as each other. Both glasses contain the same amount of apple juice.

Do you agree?

Here is another glass C, taller and thinner than the glass A or B. It is empty.

Now the apple juice from glass B is poured into the tall, thin glass C.



[*This to be done in reality, or on a video / computer*]

Look at the apple juice left in glass A, and the apple juice now in the tall glass C.
Remember, we started with the same amounts in glasses A and B. Then we poured all the juice from B to the thin glass C. Is there now:

More juice in C than A, or
More juice in A than C, or
The same amount of juice in A and C?

What makes you think so?

If you were offered glass A or C to drink, which one would you choose? Why?

Figure 1.1
Testing volume conservation

Below we will consider how items such as this may be administered. Here we will focus further on what sort of reasoning it is that we should be trying to assess. The criteria we will propose within the context of this chapter on science reasoning are that the matter to be assessed should relate to *science*, but should also relate to *general reasoning*. Furthermore it should be appropriate for children aged 6 to 12 years. The categories of reasoning from previous sections of this chapter which fit these criteria are what we described there as the thinking abilities or schemata of concrete operations and some of the schemata of formal operations. Specifically, we would include the following operations:

(1) conservations including number, matter (mass), weight, volume of liquid and displaced volume;
(2) seriation including putting things in order by one variable then re-ordering by a second variable and interpolating new objects into a series;
(3) classification including simple grouping, grouping by two variables, 'missing' groups, overlapping classes and hierarchies;
(4) cause and effect including more than one cause of one effect and more than one effect of one cause, the distinction from simple correlation, but not weighting multiple causes or probabilities; including finding simple qualitative relationships between variables;
(5) combinatorial thinking and finding combinations of up to three (or four?) variables each with two or three values;
(6) understanding a basic conception of probability and distinguishing events with lower or higher probability;
(7) basic correlative reasoning, the ability to recognise the correlation based on the proportion of events strengthening and weakening the relationship;
(8) spatial perception including perspective and mental rotation;
(9) speed in terms of distance and time;
(10) control of variables in three variable situations where each variable is directly observable;
(11) ratios of small whole numbers.

Forms of Assessing Reasoning Abilities

As indicated earlier, items assessing scientific reasoning need to be as free as possible from demands for scientific knowledge, and all required knowledge should be provided. The exercise of these aspects of scientific reasoning often requires that each item presents a series of scenarios with the response of the student at each step being observed. This approach is closely related to the principle of dynamic assessment (Tzuriel, 1998) in which what is observed is the subject's ability to learn from experience rather than their crystallised knowledge. There is a similar situation in the assessment of dynamic problem solving (Greiff & Funke, 2010), when students interact with a system presented by a computer, observe the behaviour of the system, generalise the observed rules, and then use this knowledge to solve the given problem. A similar interaction may help to activate students' thinking that then may be recorded by a computer.

For a long time, this type of testing could most reliably be managed by an individual interview and this is the basis of Piaget's clinical method. But such an interview is not a very practical approach for a classroom teacher who wishes to assess her children's current reasoning capability, nor for an education authority interested in school, regional, or national norms. In scaling up a testing method from the one-on-one assessment by a psychologist to a classroom test that can be administered by a non-specialist, some compromises of validity are inevitable. On the other hand, computerised testing can be much closer to the ideal individual interview than a paper-and-pencil assessment. Furthermore, administering the same test to every subject improves the objectivity of the assessment.

One successful example of the development of classroom tasks for assessing levels of cognitive development was the *Science Reasoning Tasks* of Shayer et al. in the 1970s (Shayer, 1970; Shayer, Adey, & Wylam, 1981). Most of the tasks developed were aimed at assessing formal operations (control and exclusion of variables, equilibrium, probability, combinations) but two were targeted at younger students:

(1) *Volume and heaviness* covers simple volume conservation up to density concepts in the Piagetian range from early concrete operations to early formal operations. The administrator demonstrates various actions (pouring liquids, lowering a mass into water in a measuring cylinder, etc.) and takes the class through the items one by one, explaining as necessary. Students answer on a sheet requiring multiple choice or short written answers. This task is suitable for students aged from 8 years upwards.

(2) *Spatial perception* is a drawing task. In one set of items students are required to predict the level of water in a jar as it is tilted (actual jars with water being demonstrated) and in others they are invited to draw a mountain, with a house on the side, then a chimney, then smoke from the chimney, also an avenue of trees going away. This task covers the range from early pre-operational to mature concrete operations and can be used with children as young as 5 years.

Even these assessment tasks are open to errors in administration and they do require some particular pieces of equipment for demonstration. The best promise for the future of assessment of reasoning including science reasoning, is the administration of tasks similar to those described above but using a computer to present the situations, to ask the questions, and even to modify the progress of the test in the light of an individual student's responses by applying the principles of adaptive testing. This approach begins to become possible when all students in a class have access to computers. As handling computers is getting easier and simpler, this promise may be realised soon. We will outline what one such test task might look like on Figure 1.2, taking the schema of classification as an example.

1 The first item screen presents an array of 4 green squares and 4 green triangles of similar size.

The instruction, delivered as text and by audio, is "Can you sort out these shapes? Drag them into two groups so all the shapes in each group are the same".

2 Array of green squares, triangles, circles.

"Divide these into three groups of similar objects."

3a A mixture of green and red squares, green and red triangles.

"Make two groups so all the shapes in each group have at least one thing in common. What feature have you used to make your groups? Colour / Shape / Size / other?"

3b "Mix them up again and then divide them into two groups in a different way. What feature have you used this time? Colour / Shape / Size / other?"

3c "Now divide them into four groups. What are the features of the shapes in each group?"

…

Figure 1.2
Classification task

Items can be added of increasing difficulty by increasing the number of variables, the number of values of each variable, by introducing empty sets (e.g., an array of red circles, red squares, blue circles), by introducing hierarchical classification, and by moving to real-life examples (e.g., farm animals). The programme would record the student's answers, assess competence in classifying at each level, offer more difficult items following success or simpler items following repeated failure, and yield an overall level of performance.

It should be possible to develop tests of this sort for each of the schemata. The question then arises, 'could just one test be developed which tested levels in all or many of the schemata?' One might have, for example, four items relating to classification, another four to conservation, more to do with causality and so on.

There are a number of reasons why such an approach may cause problems. Firstly, within each schema there are many levels of access which cannot be sampled adequately with three or four items. Secondly, in line with the relationship of this type of test with dynamic assessment, it takes a little time for subjects to 'tune in' to the topic of the test. To continually jump from one schema to another is liable to lead to an underestimation of a child's true ability as they have to 're-tune' to each new short set of questions. Finally, although the developmental progress through each schema can be mapped on to and is underpinned by a common scale of cognitive development, and one might expect a child to progress through each of the schemata more or less in synchrony, in fact, variations in experience lead to what Piaget called *decalage* – progress through one schema not keeping precisely in step with others.

For diagnostic purposes it is useful to have a profile of a child's developmental level separately in each of the aspects of science reasoning. This requires a large number of specifically prepared individual tasks. If students are systematically and regularly assessed by computer, and the results of the previous assessments are available before every testing session, the assessment may be customised for the actual developmental level of each student.

Interpretation of Assessments, Results, Strengths and Risks of Schemata Tests

Tests of science reasoning can yield valuable information at various levels. For an individual teacher, to see at first hand the responses of her pupils to a reasoning task can be quite surprising and enlightening and often elicits responses such as "I can't believe they got that 'wrong'" or

"But I only taught them that two weeks ago". Such reactions may be attributed to the fact that the nature of cognitive development and the relationship of teaching to development are often poorly understood by teachers and the results of reasoning tests can reveal that the development of reasoning such as control of variables or proportional thinking is slower than one might think, and is not amenable to simple direct instruction. Certainly, teachers can help students develop this reasoning but it is a slow process of cognitive stimulation in various contexts rather than a matter of simple instruction alone.

Once they overcome the urge to 'teach' the reasoning skills directly, teachers will find the results of reasoning tests useful to inform them of where children are now so that they can (a) map out the long road of cognitive stimulation ahead and (b) better judge what type of activities are likely to cause useful cognitive conflict – both for a class as a whole and for individual children.

On a larger scale, some national (Shayer, Küchemann, & Wylam, 1976; Shayer & Wylam, 1978) and international (Shayer, Demetriou, & Pervez, 1988) norms have been established for the ages of attainment of various levels of development which could allow a teacher, school, or education authority to make some judgement about the performance of their students compared with a wider context. Unfortunately, many of these norms are now quite old and it has been shown that the norms for, for example, the Volume and Heaviness task describe above have changed radically since they were first established in the 1970s (Shayer & Ginsburg, 2009). In spite of this shift, both by internal comparisons within a school and simply by reference to the transparent success criteria that these tests display, it would be possible even from localised testing to identify individual students who may appear to have some science reasoning disability, as well as exceptional students who might benefit from higher-level stimulation than is provided by the regular school curriculum.

The advantage of the type of test that has been discussed in this chapter is that it assesses something more fundamental than science knowledge or understanding. What is assessed has a strong developmental component, is an indicator of general reasoning ability, and underlies all effective learning. By improving the quality of assessment of science reasoning we gain a deeper insight into how our students are thinking scientifically and so are better able to help them through targeted cognitive stimulation to develop their thinking further and so provide them with the tools they need to improve all of their science learning.

But there are some features of science reasoning tests which need attention if their main purpose is not to be thwarted. Firstly, there is a small risk that some people might interpret the score from a reasoning test as a more or less fixed property of the child. Guidance on the use of the tests needs to make clear that even if the reasoning being tested is not easily amenable to direct instruction, it certainly is amenable to longer term, developmentally conscious teaching. It should be emphasised that the purpose of such a testing is to identify the need for intervention and to monitor the effects of the treatment. Science reasoning tests can be used in a formative way as well as can science knowledge tests. Furthermore, it is essential in computerised testing to apply realistic situations. Students should feel the objects and processes presented on the screen as real, otherwise they cannot make a correspondence between the real world and the one presented by the computer.

Secondly, there is the issue of test development through drafting, trialling and item statistics; re-drafting and programming the instruments for computer delivery. As indicated previously we see these tests being best administered one-on-one by individual computers. This is essentially a technical problem.

Finally, there is an issue about security, especially in systems with high-stakes testing. If the developed tests were to become freely available, and if the diagnostic purpose of the tests was misunderstood, they would be prone to coaching. That is, a school or teacher who obtained the tests and thought that there was some merit in being able to report that their students scored highly on the tests (for example in a prospectus to parents) could relatively easily coach students with ‘correct’ answers. This process short-circuits real developmental growth and the artificially inflated scores would not reflect genuine internalisation of the schemata by the students. The best guard against such misuse is education of teachers and school principals, and a policy of discouraging the public reporting of test scores of individuals or groups. The temptation of ‘teaching for testing’ or ‘test coaching’ may be further reduced if testing is regularly repeated, and the data are longitudinally connected. Artificially raising the results at one assessment point would decrease the possibility of having a gain in the consecutive assessments. Furthermore, in the case of longitudinally connected developmental data, manipulation of results may be more easily identified with statistical methods.

This raises also the issue of how such test results should be reported to students themselves. As is normal good formative assessment practice (Black, Harrison, Lee, Marshall, & Wiliam, 2003), feedback should be qualitative rather than quantitative. Simply giving a student a total score on a reasoning task is meaningless since it does not tell him or her sort of thinking at which s/he has been successful, and the sort of thinking that still needs to be developed. An efficient formative feedback, first of all, should advise students to find activities which help them to further develop and to improve the results. The test scores can only help them to control if their work has been efficient, and how it has increased since the last assessment. In a classroom, setting group feedback can actually become a teaching opportunity, as different students are invited to report their choices of answers and to justify them and engage in social construction with others.

Summary

In this chapter we have made a clear distinction between science knowledge and science reasoning, this distinction being partly clarified by their positions on a Learning-Development spectrum, which has implications for their degree of generality. As a consequence of this distinction, we also have to distinguish direct teaching and systematic stimulation of the development. This latter one is the process of improving scientific reasoning as well as fostering thinking in general.

We have seen some ways in which science reasoning may be classified and have paid particular attention to the set of scientific reasoning patterns or schemata which underpin all science learning and understanding. Science reasoning is seen as one aspect of general reasoning, or general intelligence, and both general and science reasoning are open to development through appropriate educational experiences.

We have described the nature of cognitively stimulating experience as typically involving cognitive conflict that challenges students' actual knowledge and motivates them to step further towards a higher level of understanding. We have highlighted the importance of social construction, the processes in which students dispute and argue over science phenomena mutually inspiring each others' reasoning processes. Furthermore, we emphasised the role of metacognition and the significance of becoming a conscious thinker being able to control and monitor our own reasoning processes. We have demonstrated the unique opportunities science education may offer to exercise all these essential cognitive processes.

Finally, methods of assessing students' powers of reasoning in science have been introduced. We also have provided some pointers and criteria from which it might be possible to start to develop banks of appropriate test items. The uses and potential misuses of such tests have been considered.

Formative and diagnostic assessment of scientific reasoning has already been explored in experimental educational programs for several decades. However, the demands of human and instrumental resources required for the assessment of students' reasoning prevented these methods from being broadly applied in everyday educational practice. Technology-based assessment makes personalised testing accessible in average classrooms and in this way helps to take a further step towards adjusting science education to the actual developmental level and individual needs of students.

References

Adey, P. (1998). Thinking Science: Science as a gateway to general thinking ability. In J. M. Hamers, J. E. Van Luit, & B. Csapó, (Eds.), *Teaching and learning thinking skills,* (pp. 63–80). Lisse: Swets and Zeitlinger.

Adey, P., Csapó, B., Demetriou, 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., Robertson, A., & Venville, G. (2001). *Let's Think!* Slough, UK: NFER-Nelson.

Adey, P., Robertson, A., & Venville, G. (2002). Effects of a cognitive stimulation programme on Year 1 pupils. *British Journal of Educational Psychology, 72*, 1–25.

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.
Bao, L., Cai, T., Koenig, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang Y., Li, L., & Wu, N. (2009). Learning and scientific reasoning. *Science*, 323(5914), 586–587.
Bán, S. (1998). Gondolkodás a bizonytalanról: valószínűségi és korrelatív gondolkodás [Thinking about the uncertain: probabilistic and correlational reasoning]. In B. Csapó (Ed.), *Az iskolai tudás* (pp. 221–250). Budapest: Osiris Kiadó.
Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham: Open University Press.
Boyera, T. W., Levinea, S. C. & Huttenlochera, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490.
Calvert, B. (1986). *Non-verbal test.* Windsor: NFER-Wilson.
Caroll, J. B. (1993). Human cognitive abilities. A survey of factor-analytic studies. Cambridge: Cambridge University Press.
Csapó B. (1988). *A kombinatív képesség struktúrája és fejlődése*. [Structure and development of combinative ability] Budapest: Akadémiai Kiadó.
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. (2001a). Az induktív gondolkodás fejlődésének elemzése országos reprezentatív minták alapján. [An analysis of the development of inductive reasoning on the basis of a large-scale survey] *Magyar Pedagógia*, *101*(3), 373–391.
Csapó, B. (2001b). A kombinatív képesség fejlődésének elemzése országos reprezentatív felmérés alapján. [An analysis of the development of combinative ability: a large-scale survey] *Magyar Pedagógia, 101*(4), 511–530.
Csapó B. (2003). *A képességek fejlődése és iskolai fejlesztése.* [The development of abilities and their improvement in a school context] Budapest: Akadémiai Kiadó.
Csapó, B. & Nikolov, M. (2009). The cognitive contribution to the development of proficiency in a foreign language. *Learning and Individual Differences*, *19*(2) 209–218.
Csíkos, Cs. (2007). *Metakogníció – A tudásra vonatkozó tudás pedagógiája* [Metacognition – The pedagogy of knowing about knowledge]. Budapest: Műszaki Kiadó.
Davies, J. & Graff, M. (2006). Wholist–analytic cognitive style: A matter of reflection. *Personality and Individual Differences*, *41*(6), 989–997.
Demetriou, A. (2004). Mind, intelligence, and development: A general cognitive, differential, and developmental theory of the mind. In Demetriou, A., & Raftopoulos, A. (Eds.), *Developmental change: Theories, models and measurement* (pp. 21–73). Cambridge: Cambridge University Press.
Ennis, R. H. (1995). *Critical thinking.* Prentice Hall, New York.
G. Havas, K., Demeter, K. & Falus, K. (Eds.). (1998). *Gyermekfilozófia Szöveggyűjtemény I.* Budapest: Korona Nova Kiadó.
Girotto, V. & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*(1), 325–344.

Glenda, A. (1996 August). Assessment of relational reasoning in children aged 4 to 8 years. Paper presented at the Biennial Meeting of the International Society for the Study of Behavioral Development. Quebec, Canada, August 12–16, 1996.

Glynn, S. M., Yeany, R. H., & Britton, B. K. (Eds.). (1991). *The psychology of learning science*. Hillsdale, NJ: Erlbaum.

Greiff, S. & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme In. Klieme, E., Leutner, D., & Kenk, M. (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. (pp. 216–227). 56. Beiheft der Zeitschrift für Pädagogik, Beltz: Weinheim u.a.

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.

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

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

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.

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

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

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

Klauer, K. J. (1989b). *Denktraining für Kinder I.* Göttingen: Hogrefe.

Klauer, K. J. (1996). Teaching inductive reasoning. Some theory and three experimental studies. *Learning and Instruction*, *6*(1), 37–57.

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.

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.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*(2), 155–178.

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

Lipman, M., Sharp, M., & Oscanyan, F. (1980). *Philosophy in the classroom.* (2nd ed.). Philadelphia: Temple University Press.

Lawson, A., E. (1982). The relative responsiveness of concrete operational seventh grade and college students to science instruction. *Journal of Science Teaching*, *19*(1), 63–77.

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.

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

Molnár, É. (2002). Önszabályozó tanulás: nemzetközi kutatási irányzatok és tendenciák [Self regulation: international research directions and tendencies]. *Magyar Pedagógia*, *102*(1), 63–79.

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

Nagy, J. (1987). *A rendszerezési képesség kialakulása* [Development of systematizing ability]. Budapest: Akadémiai Kiadó.

Nagy, J. (2004). A elemi kombinatív képesség kialakulásának kritériumorientált diagnosztikus feltárása [Criterion referenced diagnostic assessment of basic combinatorial skills] *Iskolakultúra, 14*(8). 3–20.

Nagy, L. (2006). *Az analógiás gondolkodás fejlesztése* [The development of analogical reasoning]. Budapest: Műszaki Könyvkiadó.

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

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

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463–466.

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities*. London: Routledge and Kegan Paul.

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

Piaget, J., & Inhelder, B. (1976). *The child's conception of space*. London: Routledge and Kegan Paul.

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

Popper, K. R. (1972). *Objective knowledge. An evolutionary approach.* Oxford: Clarendon.

Raven, J. C. (1960). *Guide to the Standard Progressive Matrices set A, B, C, D, E.* London: H. K. Lewis.

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.

Sanz de Acedo Lizarraga, M. L., Sanz de Acedo Baquedano, M. T. & Oliver, M. S. (2010). Psychological intervention in thinking skills with primary education students. *School Psychology International, 31*(2), 131–145.

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.

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

Shayer, M. (1999). Cognitive Acceleration through Science Education II: Its effect and scope. *International Journal of Science Education, 21*(8), 883–902.

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

Shayer, M., Adey, P., & Wylam, H. (1981). Group tests of cognitive development – Ideals and a realisation. *Journal of Research in Science Teaching, 18*(2), 157–168.

Shayer, M., & Adhami, M. (2011). Realising the Cognitive Potential of Children 5 to 7 with a Mathematics focus: Post-test and long-term Effects of a two-year intervention. *British Journal of Educational Psychology, in press*.

Shayer, M., Demetriou, A., & Pervez, M. (1988). The structure and scaling of concrete operational thought: Three studies in four countries. *Genetic, Social and General Psychological Monographs, 114,* 309–375.

Shayer, M., & Ginsburg, D. (2009). Thirty years on – a large anti-Flynn effect? (II): 13- and 14-year olds. Piagetian tests of formal operations norms 1976-2006/7. *British Journal of Educational Psychology, 79*(3), 409–418.

Shayer, M., Küchemann, D., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology, 46*(2), 164–173.

Shayer, M., & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children. II. 14- to 16-year-olds and sex differentials. *British Journal of Educational Psychology, 48*(1), 62–70.

Topping, K. J., & Trickey, S. (2007a). Collaborative philosophical enquiry for school children: Cognitive effects at 10–12 years. *British Journal of Educational Psychology, 77*(2), 271–278.

Topping, K. J., & Trickey, S. (2007b). Collaborative philosophical enquiry for school children: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology, 77*(4), 787–796.

Tryphon, A., & Vonèche, J. (1996). *Piaget-Vygotsky. The Social Genesis of Thought*. Hove, UK: Psychology Press (Erlbaum).

Tzuriel, D. (1998). Dynamic assessment of preschool children: characteristics and measures. In J. M. Martínez, J. Lebeer & R. Garbo (Eds.), *Is intelligence modifiable?* (pp. 95–114). Madrid: Bruño.

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.

Vidákovich, T. (1998). Tudományos és hétköznapi logika: a tanulók deduktív gondolkodása [Scientific and everyday logic: Students' deductive reasoning]. In B. Csapó (Ed.). *Az iskolai tudás* [School knowledge]. (pp. 191-220) Budapest: Osiris Kiadó.

Vygotsky, L. S. (1978). *Mind in Society*. Cambridge, Mass: Harvard University Press.

Zátonyi, S. (2001). *Képességfejlesztő fizikatanítás*. [Developing abilities through physics teaching] Budapest: Nemzeti Tankönyvkiadó.