

An example should be just an example. Which tasks really check skills?

WOJCIECH GRAJKOWSKI, BARBARA OSTROWSKA

Science Section, Educational Research Institute*

The new core curriculum for natural sciences (biology, chemistry, physics and geography) at lower secondary school level details each topic and the teaching goals involved, defining the generic skills that a lower secondary school graduate should possess. The aim of the lower secondary school leaving examination is, therefore, not only to assess memorisation of specific material but also mastery of more universal skills. The authors of the article advocate the use of test questions that check the skills included in the core curriculum and employ unfamiliar examples. In this way, it should be possible to distinguish between cases in which the student *knows the example* (remembers the facts), and the one in which they *know the principle* (have mastered the skill and know how to apply it in different situations). In addition, an analysis of example tasks of this type and the results obtained by students solving them, invites the thesis that embedding a task in an unusual context does not lead to significant increase in its difficulty.

KEYWORDS: skills assessment, exam, national curriculum, science education, test question.

This article concerns the problem of constructing tasks for school testing to assess mastery of specific skills (compared with tasks that mainly check memorisation of information). We shall start with a short discussion of the way in which the issue of development and assessment of skills is included in the new core curriculum and draw

attention to the worldwide trends with which the Polish educational reform complies. We will then look at the technical aspects of skill assessment with the use of increasingly popular multiple-choice tests, concentrating mostly on the importance of the context in which a task is embedded. Finally, we will discuss several example biology and

The article was created on the basis of the surveys: *Diagnoza kompetencji gimnazjalistów 2011* (Diagnosis of the competences of lower secondary school students 2011), *Laboratorium myślenia* (Laboratory of thinking) and *Podstawa programowa i rozwój dydaktyk przedmiotowych w opiniach nauczycieli, dyrektorów szkół oraz uczniów* (Core curriculum and development of teaching specific subjects according to teachers, principals and students) carried out within the systemic project "Quality and effectiveness of education – strengthening

of institutional research capabilities" executed by the Educational Research Institute and co-financed from the European Social Fund (Human Capital Operational Programme 2007–2013, Priority III High quality of the education system). This article was published primarily in Polish language in *Edukacja*, 119(3) 2012.

* Mail address: Wojciech Grajkowski, Pracownia Przedmiotów Przyrodniczych, Instytut Badań Edukacyjnych, ul. Górczewska 8, 01-180 Warszawa, Poland. E-mail: w.grajkowski@ibe.edu.pl

chemistry tasks (originating from the lower secondary school exam papers and used in the surveys *Laboratorium myślenia* (Laboratory of thinking) and *Diagnoza kompetencji gimnazjalistów 2011* (Diagnosis of the competences of lower secondary school students 2011) carried out by the Educational Research Institute, in which a context unknown to the student is an intentional measure to facilitate reliable verification of the mastery of a specific skill.

The authors of the article focus on discussing tasks in natural science subjects but the conclusions and recommendations formulated in the conclusions section are of a more general nature, especially if one considers the fact that core curricula of almost all subjects taught in the reformed school system strongly emphasise not only the need to teach students facts but also to develop specific skills.

New core curriculum and development of skills

The current general education core curriculum (MEN, 2009) is written in a language of requirements, which means that it directly expresses the expected educational outcomes. Its authors make us understand that what the student can do after graduation is more important than what the student does at school. Nevertheless, the current core curriculum in greater detail than in the previous one describes the teaching contents (detailed requirements) which must be incorporated. The contents are subordinate to the educational goals (general requirements), referring to skills such as drawing conclusions, identification of causal relations, or analysis of information, thus skills are of a very general nature and sometimes cross-curricular.

For proper interpretation of the new core curriculum and its efficient implementation, it is essential to understand the interrelationships between the teaching contents

(concerning what a lesson is *about*) and the implemented educational goals (determining what a lesson is *for*). For instance: the detailed requirements of the geography core curriculum, point 10.3, contains the following provision: “The student analyses graphs and numerical data concerning the development of population and urbanisation in China; explains, based on thematic maps, the differentiation of population distribution on the Chinese territory” (MEN, 2009, p. 133). At the contents level, the provision refers to a specific country, at the skills level, the provision is of a much more general nature. The student should not simply memorise specific facts concerning the development of population, urbanisation and population distribution in China, but be able to draw the right conclusions on the basis of graphs, numerical data and maps. This specific requirement is, therefore, consistent with point I of the general requirements for geography, which states that a lower secondary school graduate “can use plans, maps, photographs, drawings, graphs, statistical data, source texts” (MEN, 2009, p. 129). Thus, it should be expected, that a student who can e.g. read the value of the demographic growth for China from a statistical yearbook, will be able to do the same in the case of Iceland or Venezuela, although these states are not mentioned in the core curriculum.

A similar role is played by the compulsory experiments described in the core curricula for natural science subjects. For instance: when studying biology, the students should, “plan and carry out experiments to demonstrate that yeasts give off carbon dioxide during the process of fermentation”, or “check the impact of a selected factor on sprouting seeds” (MEN, 2009, p. 65). The aim of carrying out those experiments is not only to learn about the above-mentioned phenomena, but also, or maybe above all – to develop the skills described in point II of the educational goals

– “Knowledge of the methodology of biological research” (MEN, 2009, p. 58). Here, it is also important that *knowledge* should be of a universal nature. It means that the student who, in accordance with the provisions of the core curriculum, has learned “to plan, carry out and document simple biological experiments, to define the conditions of experiment, to differentiate between the control and research samples, and to formulate conclusions” (MEN, 2009, p. 58), will be able to apply that knowledge to every simple biological experiment, not just to those concerning fermentation or sprouting of seeds – directly mentioned in the core curriculum.

This is not, of course, a novel idea. Already in the previous examination standards (CKE, 2006), it was clearly noted that the student should be expected not only to memorise specific information but also search for and apply it (standard B II), identify and describe relationships and interdependencies between facts (standard B III) or use the knowledge obtained in problem-solving (standard B IV). According to the new core curriculum, the assumptions are repeated and emphasised by introduction of educational goals.

Skills needed in the 21st century

Stronger emphasis placed on the role of acquiring new skills in the learning process is consistent with the tendencies observed worldwide. Increasingly becomes the view that transmission of the knowledge of facts is not in itself the basic goal of education, but rather the foundation for development of skills which are initially simple, then progressively more complex. The approach is based on theoretical models, which describe learning as development and acquisition of new competences. An example of such a model is offered by the so-called SOLO taxonomy (Structure of Observed Learning Outcomes; Bigg and Collis, 1982). It differentiates between five levels of competence:

- pre-structural level – the student lacks the competence, cannot address the issue in any way;
- uni-structural level – the student is able to address a single aspect of the issue, e.g. call it or recognise a single element;
- multi-structural level – the student presents several relevant aspects (lists the elements, discusses them one by one), but is not able to describe the relations between them;

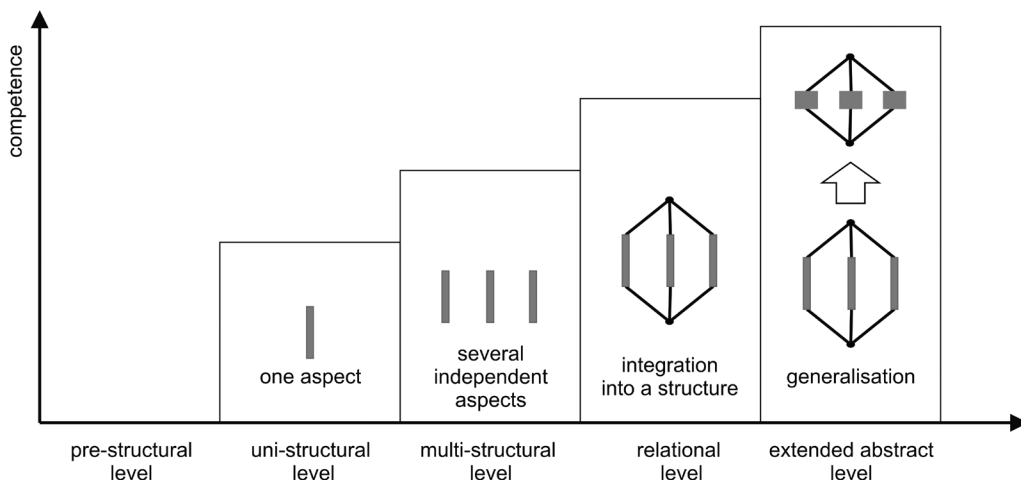


Figure 1. Competence levels according to the SOLO taxonomy (based on: <http://www.johnbiggs.com.au/academic/solo-taxonomy/>).

- relational level – the student describes the relations existing between specific elements: analyses, compares, identifies causal relations;
- extended abstract level – the student is able to generalise, i.e. apply the acquired knowledge to a new area: reason through analogy, present hypotheses, reflect and act creatively.

It should be noted that, further to the SOLO model, the growth of competences (marked conventionally on the vertical axis of the graph in Figure 1) is of a quantitative, rather than qualitative nature. In this interpretation, a student who can name the capital cities of all European countries has not reached a higher level of competence than his or her peer who knows only the capital cities of the countries neighbouring Poland. In the same way as a transition from knowledge of two features distinguishing reptiles from amphibians to the case when the student knows, say, five such features, competence cannot be said to have grown since is still related to the multi-structural level. Only the ability to analyse the meaning of those amphibian and reptile features in the context of adaptation to the environment infers reaching the relational level. The extended abstract level is achieved when a student can describe and explain those kinds of differences between various groups of organisms, that is reason by analogy.

Many other theoretical models also present acquiring competences by students in a hierarchical manner, although both their classification and the adopted terminology may differ greatly. The taxonomy of Bloom (1956) widely known in the USA and still creatively developed identifies six categories in the cognitive domain: knowledge, comprehension, application, analysis, synthesis and evaluation. In later modifications of Bloom's model, the order of categories was changed or new ones added (e.g. substituting "synthesis" with "creation"). In each variant

however, knowledge of facts (memorising) is the basis, then comprehension, and next more complex operations like analysis or evaluation.

Regardless of what term we decide to use to define those "higher skills", their significance in the contemporary world is growing. In the era of the information revolution, when one click of the mouse is enough to find out what city is the capital of any country in the world or find basic information about every Polish reptile, excessive focus on transmission of knowledge of facts is becoming pointless. Instead, more and more importance is attached to the development of skills, including higher-order skills, also called complex skills, those in the above-mentioned models that correspond to the highest levels or categories. Thinking or reasoning skills are also increasingly mentioned, although it seems that the concept is becoming slightly overused due to the high freedom of its interpretation. Also general or generic skills are defined separately, that is skills not related to a specific school subject. They may include: the ability to associate facts, draw conclusions, obtain information, learn independently or cooperate in a team. The skills are currently perceived as essential, as in the conditions of the "new economy" of a post-industrial society an employee no longer acquires an occupation for life but must be ready to retrain, at least partially, and continuously enhance his or her competences (Kearns, 2001). The school, which equips the student with a set of generic skills understood in this way, gives him or her much greater chances of coping successfully with the rapidly evolving social reality, unlike that which conveys mostly knowledge of facts.

An attempt at formation of the skills is now clearly defined in the educational documents of the highest order. In the Key Competences for Lifelong Learning (European Communities, 2007), which is an annex to a recommendation of the European

Parliament and Council of 18 December 2006, the characteristic components are defined for each key competence: knowledge, skills and attitudes. A similar definition of competences as a combination of knowledge, skills, attitudes and values was adopted in OECD's DeSeCo programme (OECD, 2005). Stronger emphasis placed on the needs to develop skills, also the complex ones, can be also perceived at the level of educational policy of specific countries. An interesting study of the ways and consequences of introducing changes to support the development of reasoning skills in three very different education systems (Northern Ireland, Israel, New Zealand) has been recently presented by the Gallagher et al. (2012). The researchers noted, among other things, the key role played by changes in the method of exam administration and assessment in the introduction of education system reforms. Evaluation of the performance of the school, whose goal is to develop students' skills, is impossible without appropriate tools enabling reliable measurement of those skills.

How to measure mastery of skills?

The examination system is a very important factor in shaping the methods of instruction. If mainly knowledge of facts is checked at external exams, it can be expected that more emphasis will be placed on memorising definitions, dates and formulas in schools. If, however, also skills specific for the subjects and cross-curricular skills are checked (e.g. reasoning, deduction, using source information, etc.), there will, with time, be a movement towards developing precisely those competences. A similar situation occurs at the level of a single class unit. If the teacher dedicates classroom time to formation of students' specific skills, but at the same time only checks knowledge of facts at tests, their students will probably focus mainly on memorising the material covered.

The method of checking student competences needs, therefore, to be adjusted to the expected teaching outcomes, which means placing more emphasis on checking the skills in Poland and many other countries. At the same time, closed tasks, that is tasks in which the student selects one of a few response options, are becoming more and more popular, both in external examinations and school-level tests. In Poland, a part of the lower secondary school leaving exam dedicated to natural science subjects is composed exclusively of such tasks. Their popularity derives mostly from ease of collection, processing and analysis of the results. Correctness of an answer can be verified by a person who is not a specialist in a given subject (with the use of the right answer sheets – or even a machine), also, doubts concerning interpretation and objective assessment of students' statements are dissipated. This obviously facilitates the operation of the examination system and lowers its costs.

Closed tasks are traditionally, although not fully justifiably, associated with checking knowledge of facts. Importantly, the popular multiple-choice questions are very often of this type, but it results not from the structural constraints, but rather market needs. As mentioned above, using closed tasks brings measurable savings. Even more can be saved (though the investment is not beneficial in the long run) by buying or creating tasks that check memorisation of names, facts and formulas; as those tasks are the fastest and easiest and simply the cheapest to create. Toch (2006) draws attention to the cheap test tasks overflowing the market, which limit checking knowledge to facts and warns of their negative impact on the quality of school instruction. Although his article discusses the situation that exists in the United States, where the entities responsible for examinations simply buy ready sets of tasks from commercial companies, we may perceive some similarities in the Polish

context, since most of the tasks that the Polish student solves during his or her education come from the collections of exercises available on the market.

It is also a widely held opinion that it is simply impossible to check the mastery of skills by means of closed tasks and they are by nature dedicated only to verification of the knowledge of facts. The publication prepared at the University of Queensland, which explains the SOLO taxonomy, offers examples of questions verifying subsequent levels of student's competence, starting from the lowest (TEDI, 2006):

- List four species of mosquito.
- List four species of mosquito commonly found in tropical areas and outline the main health risk created by each of them.
- List four species of mosquito commonly found in tropical areas and discuss their (relative) importance in public health programs.
- Discuss how you might judge the relative importance of similar threats to public health; in your discussion use various species of tropical mosquito as examples.

The very use of expressions like “discuss”, “judge” or “use as examples” in tasks checking higher-order competences suggests the need to use an open-ended task, in which the student has the opportunity to express his or her opinion in a free, longer essay.

Closed tasks, however, although they cannot substitute longer written essays or oral answers, are also useful as a way of checking skills, including those defined as “higher” or “complex”. Such tasks have been used in the OECD's international PISA (Programme for International Student Assessment) survey for years, or in the SAT and ACT tests used commonly in the United States during enrolment of candidates to institutions of higher education. In Poland, work on development of a methodology to create this type of task has been performed by experts from the Institute of Philosophy and Sociology of the Polish Academy of Sciences (Ostrowska and Spalik, 2010).

The difficult art of developing tasks

Although the use of closed tasks means facilitation and acceleration of work, their development is a difficult and labour-intensive enterprise. Determination of *what* is to be measured by means of a task should always be the starting point. In this respect, they can be divided into three categories:

- tasks checking knowledge of facts from a given subject, which require memorisation of specific information rather than mastering skills;
- tasks not requiring memorisation of knowledge, but checking general, cross-curricular skills, such as working with the text, reasoning, describing causal relations;
- tasks checking subject-specific skills, which require both the knowledge of facts from that field, and understanding of the interrelations between them and the ability to analyse and interpret them.

From the point of view of a researcher dealing with diagnosis of competences, all three types of tasks are just as valuable, as only their joint use can provide a comprehensive picture of the student's competence, with precise information on their weaknesses and strengths. In the case of school tests or external exams, where the number of tasks is limited (e.g. at the lower secondary school leaving exam, there are only six tasks per single natural science subject), tasks from the third category are particularly commendable. Their use conveys a clear message: we expect the student both to know the facts and to be able to use them. Confining oneself to tasks from the first category promotes simple learning of facts, whereas excessive use of tasks from the second category, checking only cross-curricular, generic skills, leads to a situation where student competences within a single subject are no longer checked.

The structure of a typical multiple-choice task is presented in Figure 2. The initial part is defined in the English-language literature

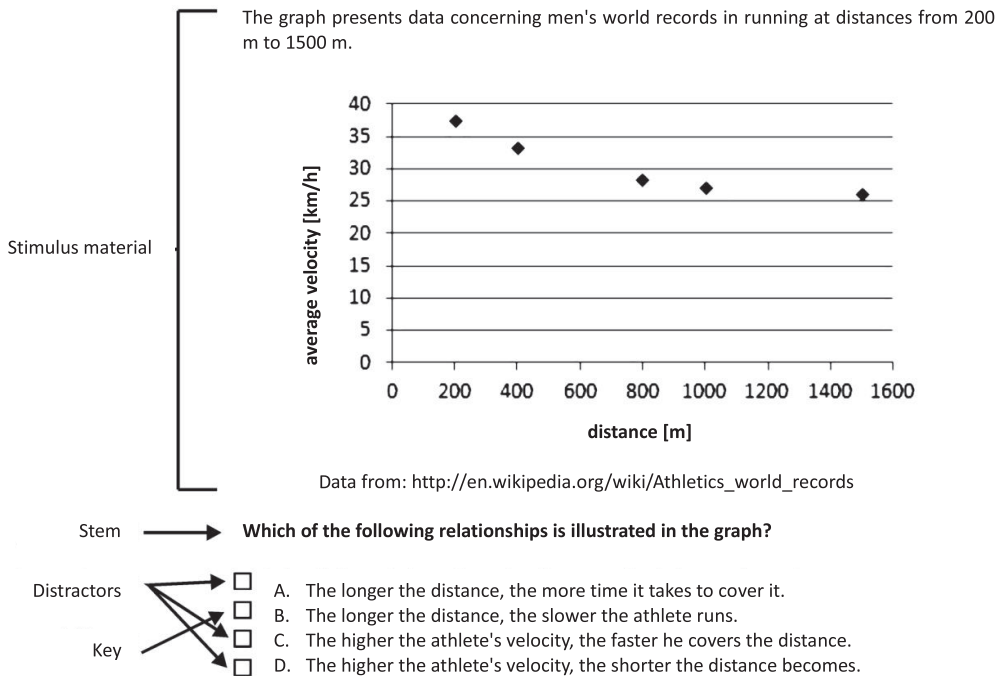


Figure 2. Structure of a typical closed task on the example of a multiple choice task. The “World records in running” task from the Teaching Tools Database of the Educational Research Institute is used (<http://bnd.ibe.edu.pl/tool-page/186>).

as “stimulus material”, while in Poland it is usually referred to as simply the “introduction”. Also the term “source material” is sometimes used, but it only refers to strictly defined information sources such as a fragment of an article, a graph or a map. Introduction, on the other hand, means everything that precedes the stem, which can include a single-sentence introduction, a representation of a chemical reaction or a short description of a real-life situation. The next part is formed by the instruction (stem), which may take an imperative or an interrogative form. At the end, there are response options, of which one is correct (key), while the others are called distractors.

In the case of tasks checking knowledge of facts, the stimulus material is usually dispensable and plays, at best, the role of embellishing the task. There is no point in placing a description of a school trip to the Tatra

Mountains there, for instance, if a traditional question about the highest peak in Poland is to be asked. If however, one wants to check skills, stimulus material is almost indispensable. In the above example, there might be included, for instance, a map of the mountain range and the student could be asked to read information concerning the highest peak from it. Thus, we arrive at the main issue to be discussed, that is the selection of the appropriate context when developing tasks for assessing skills. In our example, the Polish Tatra Mountains should not be used as the mountain range because there is a high likelihood that the student will simply remember the name, and maybe even the altitude of the peak, and it may be that, instead of assessing map skills, we will only check knowledge of the facts.

The main thesis of this article is: to check skills, it is best to use examples unknown to

the students in the tasks. We also think that, despite concerns expressed by some teachers, a strange, unfamiliar context does not constitute a serious obstacle in solving the task for the students, as long as they have mastered the relevant skill. To support our thesis, in the latter part of the article, we will present several example tasks and juxtapose teachers' opinions with the results obtained by the students solving them.

Methodology of the study

Below, we discuss closed tasks that measure skills related to natural science subjects. Two of them are from the exam paper of 2010 (CKE, 2010a). The percentage of students who correctly solved them was provided by the Central Examination Board (*Centralna Komisja Egzaminacyjna* – CKE, 2010b), and the teachers' feedback concerning one of the tasks originates from interviews carried out on commission for the Educational Research Institute (*Instytut Badań Edukacyjnych* – IBE) as part of the research programme *Podstawa programowa i rozwój dydaktyk przedmiotowych w opiniach nauczycieli, dyrektorów szkół oraz uczniów* (Core curriculum and development of teaching specific subjects according to teachers, principals and students) (IBE, forthcoming). The survey was carried out in autumn 2010 and involved performing focused group interviews (FGI) with lower secondary school natural science subject teachers¹.

The next task discussed comes from the survey *Diagnoza kompetencji gimnazjalistów 2011* carried out by the IBE in cooperation with the CKE and the regional examinations commissions in December 2011. In the survey, third grade students from 80 lower secondary schools (national, representative random sample) participated. The test

component of the survey had a form close to the lower secondary school leaving exam. The students solved tasks that checked knowledge and skills in the scope covered by the core curriculum for lower secondary school in Polish, maths, natural science subjects, history and social science. They also assessed the difficulty level of each task on a scale of 1 to 4. After completion of the field component of the survey, a meeting with 156 teachers from participating schools was held during which the tasks solved by students were discussed in subject-focused groups, the questionnaire aimed at collecting the teachers' feedback on the tasks used in the survey was distributed. Teachers were asked to assess the difficulty level of the tasks in the same way as the students. The results of the survey were put in the form of a final report (IBE, 2012a).

The last of the tasks discussed here originates from the study *Laboratorium myślenia* performed by the Science Section of the IBE. It is a longitudinal study planned for the years 2011–14, in which, by means of closed tasks, lower secondary school student skills are being assessed in the natural sciences. Each year the survey covers first grade students at a randomly selected upper secondary school who jointly solve 208 test tasks. The results of the first cycle of the survey were presented in the report (IBE, 2012b), while the statements quoted here are teachers' responses to questionnaires administered during the conference summarising the first survey cycle in Warsaw in March, 2012.

Task 1 – “Carob tree”

An example of a task that checks a specific skill, which uses an example not known to the student, is offered by task 25 from the lower secondary school leaving exam from 2010 (thus taken by students still covered by the prior core curriculum). The stimulus material describes a plant called the carob tree (CKE, 2010a, p. 9):

¹ For more information on the study refer to the IBE's website at: <http://eduentuzjasci.pl/pl/badania.html>

It is a tree of the pea family with compound, paripinnate (with an even number of leaflets) leaves. Seeds from its ripe pods – tiny, hard, with very consistent weight (197 milligrams) – were used as weights.

Four drawings of branches with fruit and leaves were placed below the description, and the student was to indicate, which of them represented the species described. It may be safely assumed that most of the students had never encountered the carob tree and were not remotely aware how it looked. However, knowledge of the plant was not necessary, the task assessed ability to recognise a species based on a description and only one of the leaves in the pictures had an even number of leaflets. Nor was the knowledge of the term “paripinnate” necessary, as the term was explained in the stimulus material.

The task was one of the easiest in the whole test – it was correctly solved by 82%, which infers that students had properly mastered the skill of interest. Nevertheless, some opinions criticising the choice of species could be heard from the teachers. Here is a statement from a teacher, coming from an interview carried out during the study *Podstawa programowa i rozwój dydaktyk przedmiotowych w opiniach nauczycieli, dyrektorów szkół oraz uczniów* (IBE, forthcoming):

This year, for instance, I do not know how the others felt about it, but in our school the biology teachers were very upset [...] there was some tree, I don't even remember what tree right now, its leaves, and the students were to tell which of pictures, and there were four of them, matched the tree, which doesn't grow in Poland at all. So we think that, in the exam, there should be for example an oak drawn. Why on earth should a child know what a tree looks like, when the tree doesn't even grow in Poland?

A clear misunderstanding can be seen here. The teacher cited thought that the task was checking knowledge of *facts* concerning

trees; while in fact, it was checking the *skill* of identification of the species based on a description. If we, in accordance with the postulate made by the teacher, placed the oak in the task, a large proportion of students would probably not need to manifest the measured skill – they would recognise the tree, as they simply know what its leaves and fruit look like. It should be noted that names of specific species almost never appear in the core curriculum, which is used to discuss specific contents and their selection is left to the authors of textbooks. Therefore, although any teacher has the right to require his or her students to know about a specific species discussed in the class, it is unacceptable that there should be questions at the external exam which would advantage students using one textbook and put students using a different one at a disadvantage. Of course, it may be presumed that only a tiny minority of students cannot recognise an oak tree, but the issue gains in importance when we consider less well known examples. The core curricula for natural science subjects contain statements like: “the student presents the basic life processes of the unicellular organisms using the example of a selected autotrophic protozoan (e.g. euglena) and a heterotrophic protozoan (e.g. paramecium)” (MEN, 2009, p. 59), or: “the student demonstrates the relationship between natural conditions and the direction and efficiency of agricultural production on the example of agriculture in France or a different European country,” (MEN, 2009, p. 133). The statements clearly indicate that only examples are mentioned and what is essential is knowledge of some general features of protozoa or agricultural countries. The tasks appearing at the lower secondary school leaving exam should, therefore, refer only to general features, in order not to put a student who learned them on the example of, say, chlorella or the Ukraine at a disadvantage. Most of all, however, it must be remembered that the goal of education is

not to provide the student with detailed information about all Polish trees or every European country. Instead, the student should be equipped with a set of skills that would enable him or her to obtain and consciously use information of that sort and those skills would be assessed in tests and examinations.

Task 2 – “Iron ores”

In the same exam paper, there is one task set in a context unfamiliar to the students, this time assessing chemistry skills (CKE, 2010a, p. 7). The task read:

Iron can be obtained from ores by reduction of the oxide using carbon. Which notation of the chemical equation is correct?

- A) $\text{Fe}_2\text{O}_3 + \text{C} \rightarrow 3\text{Fe} + \text{CO}_2$
- B) $\text{Fe}_2\text{O}_3 + 3\text{C} \rightarrow 2\text{Fe} + \text{CO}_2$
- C) $\text{Fe}_2\text{O}_3 + 2\text{C} \rightarrow 2\text{Fe} + 2\text{CO}_2$
- D) $2\text{Fe}_2\text{O}_3 + 3\text{C} \rightarrow 4\text{Fe} + 3\text{CO}_2$

Neither the past nor the current core curricula covers the process of obtaining metallic iron, nor any kind of reduction reaction. However, knowledge of those issues is not checked in the task discussed. To solve it, the simple and, significantly, generic skill of balancing chemical equations is required. Not even the formula of any chemical compound present in the equation needs to be known, they are identical in all four response options. One must understand, however, what such notation means (i.e. the numbers used). As shown by the results of most students, the use of an unusual example of a chemical equation did not prevent solution of the task and 65% of the students taking the exam gave the right answer (CKE, 2010b). If the tasks had considered e.g. burning hydrogen, the result probably would have been higher. Then, however, we would not know, if the students solving it truly demonstrated the skill of balancing chemical equations, or were simply mechanically reproducing a typical school example from memory.

Task 3 – “Neanderthal”

News of the latest achievements in the natural sciences are a commendable source of inspiration for creating tasks. Of course, the point is not to verify students' knowledge of the latest discoveries, but to check the skills described in the core curriculum on a new example, not discussed at school, which, additionally, for at least some of the students, may turn out to be interesting. Such tasks include task 4 from the natural science part of the exam paper of the *Diagnoza kompetencji gimnazjalistów 2011* (IBE, 2012a, p. 11). The task is presented below:

The Neanderthal is a fossil human, who became extinct around 25,000 years ago. Researchers have argued whether he should be considered to be the direct ancestor of modern humans. Scientists managed to read the Neanderthal genome, using his fossilised bones. The research published in 2010 suggests that 1–4% of genes in the modern non-African human genome might come from the Neanderthals. However, no such genes were found among native Africans.

Finish the sentence, selecting one of the response options provided below.

The discovery described above means that the Neanderthal:

- A) cross-bred with the ancestor of modern humans.
- B) is a direct ancestor of modern humans.
- C) had a different genetic code than modern humans.
- D) was superseded by modern humans.

The task gave rise to serious controversies among the teachers participating in the conference summarising the results of the DKG survey. They pointed out that the Neanderthal does not appear in the biology core curriculum for the lower secondary school, and many believed that the complicated text introducing new, difficult concepts constituted a serious barrier for students. In the

questionnaire aimed at collection of the feedback, statements like the following were included: “Too much of the stimulus material, many difficult words for the student. At the very beginning, he or she is discouraged from further reading the instructions and response options”; “The Neanderthal is not in the core curriculum. The student and the parents can see that in the textbook”; The very word *Neanderthal* frightened the students. The task does not check any specific knowledge”; “[The task checks – author’s note] reception of information from the media – if the students are not interested in a given topic, they will not know”. The teachers responding to the questionnaire rated the difficulty of the task at 3.5 points (biology teachers) and 3.3 points (teachers of other natural science subjects) on a scale of 1 to 4, thus considering it to be the most difficult in the part dedicated to biology (IBE, 2012a).

As a matter of fact, however, the task did not check any knowledge concerning the Neanderthal, but the ability to interpret the text based on knowledge of evolution and genetics in the scope included in the core curriculum. It should be mentioned that at the time of performing the survey (December 2011) the sections on Genetics and Evolution of Life had not been covered in most of the schools yet. Nevertheless, 44% of students solved the task correctly, which is a value close the test average (IBE, 2012a). It seems, therefore, that indeed, the ability to analyse the text was more important and enabled the student to cope, despite missing knowledge.

As regards students, they considered the task to be much easier, assigning the average of 2.6 points on a scale of 1 to 4 (IBE, 2012a). Some of the teachers also held opinions different from those quoted above, e.g. “To understand the question, it is enough to read it, knowing nothing about the Neanderthal. The question connects evolutionism and genetics. Students should be reminded to

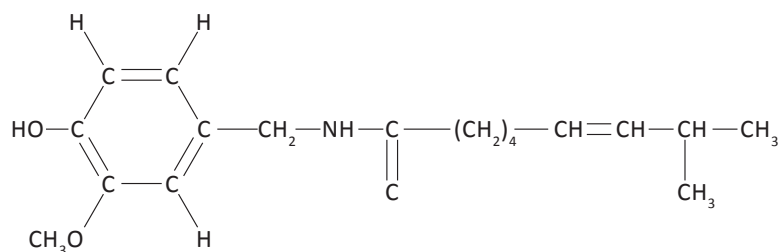
read the stimulus material”. The last sentence seems particularly meaningful, as the teachers present at both conferences (summarising the results of the surveys *Diagnoza kompetencji gimnazjalistów 2011* and *Laboratorium myślenia*) generally agreed that students solving tasks often fail to read (or just scan) not only the stimulus material, but also the instruction. Many conference participants linked it to the vanishing ability of teenagers to focus attention on a longer text, caused, at least partially, by the contemporary, shorthand and condensed transmission of information in the media. A point already mentioned should be remembered here – stimulus material encountered in tasks exclusively checking facts is dispensable. Since students are still mostly used to these, it may be presumed that they had not developed the habit of close reading of the stimulus material, as they had not needed it previously in tests. The growing popularity of tasks such as the one quoted above, where close reading of the stimulus material is necessary to provide the right answer due to the unfamiliar context, may result in students starting to pay more attention to the stimulus material and the precise meaning of the instruction.

In the case of the three examples already discussed, the results obtained by students suggested that the unusual context of the task did not significantly affect its difficulty. Finally, we would like to present a task, which proved extremely difficult, and consider to what extent the very low percentage of wrong answers resulted from the use of an example unknown to the students, and to what extent it was determined by other factors.

Task 4 – “Hot taste of chilli”

The task was used in the first cycle of the survey *Laboratorium myślenia*, and was then declassified and placed on the national Teaching Tools Database (*Baza Narzędzi Dydaktycznych* IBE, 2012b):

Capsaicin is an organic compound responsible for the hot, spicy taste of chilli peppers. The formula of the compound is presented below.



Indicate which groups characteristic for chemical compounds can be found in the capsaicin formula.

Group	Is it present?
I. carboxyl	<input type="checkbox"/> YES / <input type="checkbox"/> NO
II. ester	<input type="checkbox"/> YES / <input type="checkbox"/> NO
III. multiple bonds	<input type="checkbox"/> YES / <input type="checkbox"/> NO

The students found the task very difficult. Lack of the carboxyl group was correctly ascertained by just 33.6% of the respondents, and lack of the ester group – 49.2%. Only identification of the multiple bond (72.4% of correct answers) turned out to be relatively easy; nevertheless, the whole task was correctly solved by only 6.9% of people (IBE, 2012b).

According to the teachers participating in the conference summarising the survey results, the problem lay in the very complex chemical formula, which was unknown to the students. The questionnaire assessing the task included, among others, the following statements: “A complex formula – the student may not undertake [an attempt at solving] the task.”; “This is a task addressed at students interested in chemistry or a competition task.”; “The student will be confused with such integrated knowledge in one complex formula.” Is it really the case that the use of the capsaicin formula in the task was a bad idea? Probably the task, in which the student had been asked about the presence of the carboxyl group in the formula of e.g. acetic acid, would not have

raised so much controversy. Yet can we say that a student who can identify a particular group only in a compound known from lessons can recognise the group? Chemistry is different from other natural sciences due to the high number of symbolic notations (elements, formulas, chemical equations) and the symbols are used precisely to allow their arrangement in different ways to create new information. Someone who cannot read them does not really know the meaning of the symbols.

Often, however, as teachers underline, a difficult, unusual context of the task may effectively discourage the student from even attempting to solve it. In the case of a closed task, now used in the lower secondary school exam and in the survey *Laboratorium myślenia*, it is still difficult to detect a situation when a given person gave up solving it, based on results only. Polish students are well trained in answering test questions and they know that that even if they do not know the correct answer, it pays to make a random guess. In the case of the task discussed, in a single line, fewer than 3% of the students

checked nothing, while in the whole task – fewer than one percent (IBE, 2012b). This does not infer, however, that others tried to solve the task – some of them probably made a random guess without much thought. Therefore, it is not possible, based on the results, to determine whether the complex formula and stimulus material starting with the word “capsaicin” effectively discouraged a large proportion of the respondents or if the cause was mainly insufficient mastery of the relevant skill.

Taking into account the results obtained by students solving the three tasks described above, we incline to the belief that the problem here was failure to recognise or know the functional groups, rather than intimidation by the example used. Similarly, in the other tasks discussed above, the point was not to know about the carob tree or the iron ore reduction reaction or to know about Neanderthals. The tasks assessed the skill of recognition of a species based on description, understanding the essence of chemical equation notation and interpretation of a text on the basis of general biological knowledge.

Conclusions

In this article, we tried to prove that tasks assessing skills should refer to examples unknown to students and therefore different to those used in textbooks or traditionally discussed in the classroom. In such a case, the student cannot refer to information learned and has to demonstrate true knowledge of the relevant skill. We think that tasks constructed in this way should appear as often as possible both in classroom tests and at external examination. Since the examination system is a very effective tool to influence school teaching methods, it should be expected that placing more emphasis on assessing skills will lead to their better development. It is extremely important, though, that all interested parties in the education system,

including teachers, students and their parents, should have no doubts about what methods can be employed to assess competence and why. The statements of teachers quoted indicate that at least some misunderstood the idea of assessing skills, remarking primarily on the presence of information in the tasks not covered by the core curriculum. If teachers themselves have these kinds of doubts, it is difficult to require students and their parents to fully appreciate the justification of this method of assessing competences. Along with introduction of this type of skill-measuring task to exam papers, textbooks and collections of exercises, appropriate publicity activities should be offered by institutions responsible for teacher training. We hope that, in the future, students will not react to the occurrence of the corab tree, the Neanderthal or capsaicin with the outcry “... but it was not covered in the class!” but he or she will think “we covered it, but in a different example”. Indeed, examples discussed in the classroom constitute only a tiny fraction of the knowledge available to young people, and school should focus primarily on teaching how knowledge can be acquired and used.

Literature

- Biggs, J. B. (n.d.). SOLO taxonomy [Chart]. Retrieved from <http://www.johnbiggs.com.au/academic/solo-taxonomy/>
- Biggs, J. B. and Collis, K. F. (1982). *Evaluating the Quality of Learning – the SOLO Taxonomy*. New York: Academic Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive Domain*. New York: Longmans.
- CKE [Centralna Komisja Egzaminacyjna – Central Examination Board] (2006). *Standardy wymagań będące podstawą przeprowadzania egzaminu w ostatnim roku nauki w gimnazjum [Standards forming the basis for carrying out the test in the last year*

- in lower secondary school]. Retrieved from http://www.cke.edu.pl/images/stories/000011_gim_pr_mas_gimn.pdf
- CKE (2010a). *Egzamin w klasie trzeciej gimnazjum z zakresu przedmiotów matematyczno-przyrodniczych [The exam in third grade of lower secondary school. Mathematics and science]*. Retrieved from http://www.cke.edu.pl/images/stories/001_Gimnazjum/gm_1_102.pdf
- CKE (2010b). *Osiągnięcia uczniów kończących gimnazjum w 2010 roku [Achievement of students graduating from lower secondary school in 2010]*. Retrieved from http://www.cke.edu.pl/images/stories/001_Gimnazjum/spr_gimn_2010.pdf
- Gallagher, C., Hipkins, R. and Zohar, A. (2012). Positioning thinking within national curriculum and assessment systems: perspectives from Israel, New Zealand and Northern Ireland. *Thinking Skills and Creativity*, 7/2012, 134–143.
- IBE [Instytut Badań Edukacyjnych – Educational Research Institute]. (forthcoming). *Podstawa programowa i rozwój dydaktyk przedmiotowych w opiniach nauczycieli, dyrektorów szkół oraz uczniów [Curriculum and teaching development in the opinions of teachers, head teachers and pupils]*. Survey report. Warszawa: Instytut Badań Edukacyjnych [Educational Research Institute].
- IBE (2012a). *Diagnoza kompetencji gimnazjalistów. Przedmioty przyrodnicze [Diagnosis of the competences of lower secondary school students. Science]*. Retrieved from http://eduentuzjasci.pl/images/stories/badania/diagnizakg/dkg_przyroda.pdf
- IBE (2012b). *Laboratorium Myślenia – raport z pierwszego cyklu badań [Laboratory of thinking – the report of the 1st cycle of studies]*. Unpublished typescript. Warszawa: Instytut Badań Edukacyjnych.
- Kearns, P. (2001). *Generic Skills for the New Economy*. Leabrook, Australia: Australian National Training Authority.
- MEN [Ministerstwo Edukacji Narodowej – Ministry of National Education] (2009). *Podstawa programowa z komentarzami (t. 5: Edukacja przyrodnicza w szkole podstawowej, gimnazjum i liceum) [The core curriculum with comments (vol. 5: Environmental education in elementary, lower secondary and high school)]*. Warszawa: Ministerstwo Edukacji Narodowej.
- OECD [Organization for Economic Co-operation and Development] (2005). *The definition and selection of key competencies. executive summary*. Retrieved from www.oecd.org/dataoecd/47/61/35070367.pdf
- Ostrowska, B. and Spalik, K. (eds.). (2010). *Umiejętności złożone w nauczaniu historii i przedmiotów przyrodniczych [Complex skills in the teaching of history and science]*. Warszawa: Wydawnictwo IFiS PAN. Retrieved from <http://eduentuzjasci.pl/pliki/umiejtnosci.pdf>
- TEDI [Teaching and Educational Development Institute] (2006). *Biggs' structure of the observed learning outcome (SOLO) taxonomy*. Retrieved from http://creative.canberra.edu.au/groupwork/documents/Biggs_Solo.pdf
- Toch, T. (2006). *Margins of error: the testing industry in the No Child Left Behind era*. Washington: Education Sector. Retrieved from <http://www.educationsector.org/publications/margins-error-testing-industry-no-child-left-behind-era>
- European Communities (2007). *Kompetencje kluczowe w uczeniu się przez całe życie – europejskie ramy odniesienia [Key competences for lifelong learning - a European reference framework]*. Luxembourg: Publications Office of the European Union. Retrieved from [Wspólnoty Europejskie /dgs/education_culture/publ/pdf/ll-learning/keycomp_pl.pdf](http://Wspólnoty_Europejskie/dgs/education_culture/publ/pdf/ll-learning/keycomp_pl.pdf)

Acknowledgements

While working on this article, the authors drew extensively from the work of the Science Section of the IBE, and the main theses presented herein are to a great extent the result of discussions between its members and regular associates who participated in cyclic workshops dedicated to the development of test tasks.