# Annex 11:

**Framework for National Assessment of STEM competences in Montenegro**



**Project "Integration of key competencies into the education system of Montenegro"**

co-financed by the European Union and the Government of Montenegro, implemented by the EPRD consortium

**Version:** 31 August 2020

|  |
| --- |
| * *The contents of this publication are the sole responsibility of the EPRD Consortium and can in no way be taken to reflect the views of the European Union.* * *Sentences in this publication written in the same gender refer equally to males and females.* |

Content

[1. Introduction 5](#_Toc52194282)

[2. Curriculum 5](#_Toc52194283)

[3. Exams and assessment 6](#_Toc52194284)

[**3.1.** **National knowledge assessments** 7](#_Toc52194285)

[4. Results of international surveys 8](#_Toc52194286)

[**4.1.** **PISA test** 8](#_Toc52194287)

[5. STEM skills and attitudes 10](#_Toc52194288)

[6. What kind of testing of STEM competencies is needed by the educational system of Montenegro? 11](#_Toc52194289)

[7. Recommendations 12](#_Toc52194290)

# Introduction

This document was developed within the Project "Integration of key competencies in the education system of Montenegro" co-financed by the European Union and the Government of Montenegro through the IPA 2 Program of the EU and Montenegro for employment, education and social protection. The project is two-years long and has been implemented since 31 August 2019 to 30 August 2021 in order to improve the quality of primary and secondary education and support for initial education and continuous professional development of teachers and quality assurance at the level of primary, secondary and higher education. The document is based on the analysis of the assessment system, and national and international research in the field of STEM in primary education in Montenegro.

National knowledge tests, i.e. testing at the end of the II and III cycles, are provided by law as part of the quality assurance system. The aim of introducing these elements into the education system of Montenegro is to monitor the educational achievements and the achievement of subject outcomes for several subjects. For cross-curricular topics and the development of key competences, however, nowhere is the method of examination explicitly provided. Increased interest in the field of mathematics, natural sciences, engineering, and technology (STEM) in recent years as the basis of economic development highlights the need for a special framework for national testing of STEM competencies in Montenegro. This document analyses the main characteristics of the assessment system, national and international knowledge tests and based on them several recommendations for the introduction of STEM testing.

# Curriculum

The domain of the STEM field covers the subjects of Mathematics in all cycles, Nature and Society in the first cycle, Nature in the second cycle and Biology, Physics, Geography, Chemistry and Informatics with technology in the second and third cycle. The total number of science classes is lower than in most European countries, while the number of Mathematics classes is typical for Europe. Students have their first meeting with informatics and technology in the fifth grade within the joint subject Informatics with technology with a very modest fund of classes - one hour a week.

***Table 1: Total number of hours of STEM subjects per class***

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subjects:** | **I** | **II** | **III** | **IV** | **V** | **VI** | **VII** | **VIII** | **IX** |
| Mathematics | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Nature and society | 2 | 2 | 2 |  |  |  |  |  |  |
| The nature |  |  |  | 2 | 2 |  |  |  |  |
| Biology |  |  |  |  |  | 2 | 2 | 1 | 1 |
| Physics |  |  |  |  |  |  | 1 | 2 | 2 |
| Geography |  |  |  |  |  |  | 2 | 2 | 1 |
| Chemistry |  |  |  |  |  |  | 1 | 1 | 2 |
| Informatics with technology |  |  |  |  | 1 | 1 | 1 | 1 |  |

In the educational system of Montenegro, up to the fifth grade of primary school, natural sciences are studied integrally within one subject with a fund of two hours per week. In the sixth grade, the focus of natural sciences is significantly narrowed because only Biology is studied, which represents a discontinuity in the study of physical phenomena and phenomena related to the planet Earth.

Current subject programs in primary schools were introduced in the 2017/2018 school year. The reason for the program revision was primarily the determination to introduce an outcome-based learning model. This revision of the program was also of an organizational nature, thus changing the subject programs with a reduction in the number of hours per subject. The reduction of the content of the subject programs was aimed at relieving of non-functional knowledge and skills while providing a greater individual approach to students. Today, all subjects of the STEM area are characterized by subject orientation, defined subject outcomes, as well as ways and criteria for evaluating outcomes, explicit links with other subjects and interdisciplinary topics.

This approach to learning implies a greater representation of various active forms of teaching and learning, where students participate in activities that require them to engage higher cognitive levels such as project learning, laboratory work, argumentation, critical thinking, data interpretation, formulation of conclusions, teamwork, etc. Such high goals should provide enough hours and time for extracurricular activities, which is still not the case. The lack of time for projects and thematic teaching makes some outcomes neglected while favouring factual knowledge and application based on solving classical mathematical equations.

Revision of the program from 2017/2018 supports the autonomy of teachers and schools, which is ensured through the "open part" of the program, which is represented by 15-20%. This is the untapped potential of the Montenegrin education system, which, with sufficient teacher training, would enable flexible prioritization between subject and interdisciplinary outcomes.

# Exams and assessment

Assessment of the achievement of educational goals and subject outcomes is carried out through internal (school) and external assessment. In both cases, assessment is mostly summative, although teachers are expected to use assessment for learning, i.e. formative assessment in their work. The subject outcomes, as well as the criteria for their evaluation, are formulated in such a way as to facilitate the evaluation of outcomes for the needs of school assessment. In this way, outcomes are unconsciously avoided that we cannot directly measure, which has the consequence that students in Montenegro are still expected to have primarily reproductive knowledge.

School assessment is also covered in the document Standards of Competences for Teachers and Principals of Educational Institutions, as a special standard (assessment and evaluation). The document is very informative regarding the importance and function of assessment, it refers to the key principles that should be respected in the process of student assessment, and lists different ways of assessment, which favour different ways of activating students: "Teacher continuously and effectively uses summative assessment to support, explain and document the learning process; regularly and timely provides students with feedback on learning achievements, encourages them to self-assess and set goals for their own learning; knows how to analyse data obtained on the basis of internal and external testing of knowledge, skills and understanding and uses them to improve the teaching and learning process."

## **National knowledge assessments**

In accordance with the amendments to the education laws in 2017, national knowledge tests were introduced, i.e. testing at the end of the II and III cycles. According to these regulations, at the end of the second educational cycle, the Examination Centre should test the achievements of students in Montenegrin, i.e. mother tongue, mathematics, English, or selected chapters from natural or social sciences. The selected chapters of natural sciences are not precisely defined. The practice of national knowledge tests, especially in STEM fields, so far (October 2020) has not yet taken root. After a ten-year break from the previous initiative, trial tests are being conducted again. During 2018, a trial test was done from selected chapters from natural or social sciences, in 2019 for Montenegrin-Serbian, Bosnian, Croatian language, i.e. mother tongue, while in 2020 there should have been a test in mathematics, but it was postponed to 2021 due to the pandemic. At the end of the third cycle of primary school, an external test of students' knowledge of: Montenegrin or mother tongue, mathematics and one subject that the student chooses from the list of compulsory subjects, and which are studied in the third cycle in at least two grades. The preparation of tasks is performed by the Examination Centre, and the examination of knowledge is performed by the Examination Centre in cooperation with the school.

The purposes of testing at the end of the second and at the end of the third cycle are not the same. The exam after the third cycle serves the selection for enrolment in high schools, so its function is primarily selective, while the national testing after the second cycle should have, above all, an evaluation function. The results of these tests should also be used to monitor educational reforms and correct the education system. Student achievement on these tests does not affect school grades.

For the national exam at the end of the third cycle, there are exam catalogues that define in more detail the topics and questions used in the exam. Exam catalogues also prescribe how many tasks should be from certain areas and what should be their distribution according to the difficulty and type of tasks. Despite the fact that the distribution according to these parameters is predefined (planned test specification, i.e. exam specification), in the current procedure there is no control of the validity of the tasks and specific tests do not have their own specification[[1]](#footnote-1) in which these parameters would be specified. As the tests do not have specifications, it is not possible to perform a statistical analysis of the tasks used to determine whether the requirements from the examination catalogues are implemented.

Comparability of tests is provided by the same number and type of tasks from year to year, as well as the time provided for development. Due to the small population of students in Montenegro, it is very difficult to organize pre-testing of tasks and keep them secret, and public pressure, given that the exam has a selection role, is such that it is necessary to publish the content of tests. Due to all the above, in each school year, the Examination Centre appoints Commissions for deciding on the limits for exam grades, i.e. performance assessment on individual tasks.

The tests from the national exam also lack a deeper statistical analysis that would show the statistical parameters of each item of the test as a whole. Secondary analyses of tests should be built on these results, which would put the statistical characteristics of individual tasks in the context of the tasks themselves and the circumstances in which the students were examined..

From the aspect of examining competencies in natural sciences, the choice of subjects that students do at the end of the third cycle, diminishes the importance of the test and essentially destroys the assumption of sample representativeness. Participants mostly choose to take an English language test, which is why very few students choose to take science tests. Typically, 200-300 students choose the Biology test, while less than 100 students choose the Physics test. The results on the test obtained in this way cannot have the metric characteristics needed to monitor the educational system. Bearing in mind how difficult it is to create new quality tasks, this practice of "spending" good tasks is not particularly rational.

Apart from the representativeness of the sample, another major shortcoming of the national exam is the practice according to which all the tasks that were on the test after that are published. In this way, it is impossible to compare the tests, i.e. the results that were achieved on them, and longitudinal achievement studies are not possible. In other words, we cannot say whether the achievement in one year is better or worse than in the previous year if the comparability of tests is not ensured. For comparability, it is necessary to have the same scale for reporting student achievement for different tests, as a basic part of the test that measures achievement in the same subject.

It should be noted that there are no national tests of knowledge concerning technique and technology. These topics are not included in international studies either, so there is practically no data on students' achievements in the field of technique and technology.

In any case, there are high expectations in Montenegro that national testing could be the basis for assessing the quality of educational outcomes. The results of national testing, external knowledge testing and PISA testing are also used by the education system for external evaluation of school quality.

# Results of international surveys

Montenegro has been continuously participating in the PISA study for fifteen years, so it has results and experience in their interpretation for the entire five cycles of this research. The target group of this study are fifteen-year-olds, i.e. students who have just left primary school or are, in smaller numbers, at the end of primary school. The topics of the PISA study are mathematics, science and reading as standard. Due to the topics and ages of the respondents, PISA should be an ideal study for analysing the achievements of Montenegrin students at the end of primary school in the STEM field.

In addition to the PISA study, in the field of mathematics and natural sciences, Montenegro also participates in the TIMSS 4 international study, which examines the knowledge and attitudes of fourth grade students. Students from Montenegro participated in this study for the first time in 2019 and the results that we could analyse are not yet available.

## **PISA test**

According to the results so far at PISA, it can be concluded that Montenegro is making significant progress in the field of mathematics (in 2006, the average achievement was 399 points, while in 2018 it rose to 430). In the domain of natural sciences, unfortunately, there is no such trend (Table 2). The differences from year to year are within the statistical error and we can conclude that the achievements from Montenegro in this area are stagnant.

**Table 2: Average achievement of students from Montenegro in the domains of mathematics and science at PISA from 2006 to 2018**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **2006** | **2009** | **2012** | **2015** | **2018** |
| Science | 412 | 401 | 410 | 411 | 415 |
| mathematics | 399 | 403 | 410 | 418 | 430 |

The most important results of PISA for Montenegro do not refer to the achievement itself, but to the distribution of students by levels of achievement. According to the PISA results, approximately 50% of Montenegrin students in the field of mathematics and science do not achieve the second level of achievement, which is considered the minimum acceptable level of functional literacy according to the PISA study (Table 3). We remind you that the Europe 2020 strategy aimed to reduce the percentage of students with insufficient functional literacy to 15% in European societies by 2020. Now, Montenegro still has a lot to do on that goal. Students who have not achieved at least the second level of achievement are in the zone of educational risk since their level of functional literacy threatens their progress through schooling and opportunities for choosing a profession. These are students who have just started secondary education and who still have a long period to overcome educational challenges, and their current competencies are not enough for that.

**Table 3: Percentage of students by achievement levels at PISA 2018**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Below 1st level** | **1st level** | **2nd level** | **3rd level** | **4th level** | **5th level** | **6th level** |
| Mathematics | 19.9 | 26.3 | 27.3 | 17.9 | 6.9 | 1.6 | 0.2 |
| Science | 16.8[[2]](#footnote-2) | 31.4 | 31.5 | 15.9 | 4.0 | 0.3 | 0.0 |

Bearing in mind that at the first and second level of achievement there are not many tasks of higher cognitive level, it is important to note that the PISA test in Montenegro does not effectively measure what is announced as the focus of testing but only low level achievements. Also, according to the specification of the PISA test from 2005, where the focus of the research was science, only 8% of the tasks referred to low achievements, while for medium there were 30% and high 61%. This means that we obtain information on student achievement in Montenegro based on an exceedingly small number of tasks that do not examine higher cognitive levels (Table 4).

**Table 4:** **Description of cognitive requirements by task levels**

|  |  |
| --- | --- |
| Level of knowledge | Cognitive requirements |
| Low | Performing a procedure that has only one step, recalling one fact, term, principle, or concept, or locating one piece of information from a diagram or table. |
| Medium | Using and applying conceptual knowledge to describe or explain phenomena, select appropriate procedures that include two or more steps, organize / present data, interpret, or use simple data sets or graphs. |
| High | Analyzing complex information or data, synthesizing or evaluating evidence, justifying, reasoning using different sources, devising a plan or sequence of steps to solve a problem. |

Another particularly important finding of the PISA study relates to differences between types of schools (Table 5). The difference between students who are in high school in relation to secondary vocational schools is more than 50 points, and in relation to three-year vocational schools even more than 100 points.

**Table 5: Average student achievement by type of school at PISA 2015**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **High School** | **4-year vocational school** | **3-year vocational school** |
| Mathematical literacy | 460 | 400 | 343 |
| Scientific literacy | 454 | 399 | 333 |

PISA test results are taken globally as relevant indicators of the quality and availability of skilled labour. Therefore, the finding that shows that most students in vocational schools in the field of mathematical and scientific literacy are functionally illiterate is alarming. PISA is certainly not an optimal test for examining the competencies of students in secondary vocational schools and therefore it is necessary to consider what are the planned outcomes of vocational education, and what are the real results, i.e. how to adapt the practice of assessment and external testing to vocational schools.

From the aspect of the effectiveness of measuring achievement, the differences by types of schools are so great that the question arises whether students from high schools and secondary vocational schools should take the same test.

Participation in PISA is provided by law, so the progress of the reform of the education system is monitored through the results of the PISA study. The impact of the PISA study is also reflected in the intention to use for national examinations competencies that are "related" to the competencies and definitions of literacy on the basis of which PISA achievement levels and PISA tasks have been operationalized. This should mean that the outcomes and standards in question are more like a framework for a PISA test. Judging by the theses presented in the publication "Results of the PISA 2015 study and recommendations for educational policies", the higher the similarity, the greater the harmonization of national outcomes with international ones, so schools, teachers and students send a clear signal of what is expected and evaluated as quality educational outcome. Given that there is a clear commitment of Montenegro to regularly participate in the PISA study, the procedures for preparation, implementation and analysis of the PISA test should have a greater impact on the practice of national knowledge tests in Montenegro. Despite the reports on international testing published by the Examination Centre, it is evident that other institutions (Bureau for Education Services, Centre for Vocational Education, Institute for School Textbooks) are not informed about PISA testing and relevant results, as well as the use of results for creating educational policy.

# STEM skills and attitudes

Competency testing in the STEM field at school and national level comes down to measuring knowledge and attempts to recognize skills and attitudes in competencies are exceedingly rare. International studies PISA and TIMSS pay more attention to this but assessing the level of competencies is limited to a form of test in which there is no room for demonstration exercise, practical or teamwork, creative solution, or research.

For the measurement of skills to be relevant and reliable, it is necessary to collect data on the current context of the development of these skills in Montenegrin schools. There is very little data on how much school school equipment is needed, how much time students spend doing experiments or laboratory exercises, how much time teachers devote to analysing such activities, and how students show that they have mastered these skills. There is even less data on how many students know how to use basic instruments and benchmarks, e.g. to use a stopwatch, to find the northeast with the help of a compass, or to measure the mass of an apple on a scale. The TIMSS 2019 study collected data of this type, but they are not yet available to us.

Laboratory equipment is not maintained and renewed in an adequate way in most schools, among other things due to the lack of teachers and trained technical staff, so that today there are few schools in which laboratory exercises are actually provided by the program.

Skills related to learning and applying science are no less important than the scientific knowledge acquired in schools. As the level of knowledge is much easier to measure with control exercises and oral examination than the level of skills, this has the consequence that the emphasis of assessment and thus learning is shifted to testing knowledge and application of knowledge in computational tasks. There are not enough tests in control tasks concerning students' competencies for measuring, analysing data, and situations, or constructing and verifying models. By the fact that these abilities are not evaluated, incomparably less attention is paid to them. Teachers often "sacrifice" classes to develop project-based learning skills through experiments, discussions, or research to work more on what will really be on the test. This approach has made practical, field and laboratory activities in teaching natural sciences practically meaningless in many schools.

An additional problem is created by the division of natural sciences into special, almost unrelated, and uncorrelated subjects in the older grades of primary school. Natural phenomena that would be relatively simple to present in thematic teaching, get much more formal and narrower outlines at the level of individual subjects.

At school, subjects in which natural sciences and mathematics are studied are not attractive enough and very often students consider them to be difficult subjects for which they are not sure what exactly they are used for. Most of the examples of natural phenomena that have been used in textbooks in recent decades, especially those related to physical principles, are not recognized by students today because they did not have direct experience with these phenomena. Experiments with levers, candles, or pouring liquid through a rubber tube, for example, are not part of the cognitive experience for most students. To be able to improve the situation, we first need to have data on what is known to the students from the mentioned experiments, and what is not.

# What kind of testing of STEM competencies is needed by the educational system of Montenegro?

The end of the second cycle of primary education is not the moment when the selection of students for continuing education is done, nor are certificates of successfully completed education issued. Accordingly, the only real function of the national examination could be **evaluation**, i.e. to determine what students know or do not know, can or do not know. This means that the test should cover as many outcomes as possible and the test should check the effect on as many tasks as possible. The achievement of individual students should be in the background. A much more important result should be how many students have successfully solved a certain task than how many tasks a student has solved. This test setup allows the use of parallel forms of the test, i.e. that not all students necessarily have to do the same test and the same tasks. Evaluation of student competencies can also enable evaluation of the work of schools or individual teachers. If the function of the test were extended to the evaluation of teachers 'work, it would be necessary to measure the effects of students' socio-economic status on the development of competencies, which would probably be too complex and ambitious at this time.

The aim of the national survey is primarily to verify the achievement of outcomes. The test by which the examination is performed can be criterion (if we describe the desired outcomes or standards of achievement precisely enough) or **normative** if we are primarily interested in comparing the achievements of students in the population. This choice is difficult to make before trial tests are done. It often happens that we are not able to check the achievement of many outcomes with a test, which calls into question the representativeness of the outcomes and the criteria for their achievement. Therefore, it is easier to make a new test as normative. This test should include tasks of different levels of difficulty that cover as many outcomes as possible.

In both cases, for both the criterion and the normative test, the test must be of a **summative** nature because the goal is to measure what and how much students know at the national level, not to measure knowledge to encourage individual student learning which is a characteristic of formative tests. The fact that the test would be summative does not mean that it would have to be especially important for students, i.e. to bring them grades, points, or certificates. This is exactly the case with the PISA test. It is a normative and summative low-risk test on which nothing depends on the students themselves. The results of such a test are much more significant for the education system than with individual students. However, one should also keep in mind the possibility for the national test to be graded, i.e. for it to be an external test that all schools use once a year to assess students. Thus, the practice of assessment in school would be "calibrated" by one external test that would be the same for everything that could have positive effects on the uniformity of assessment criteria. However, the national test being evaluated would carry higher risks. Greater attention would have to be paid to the security of tests and coding, and it is likely that public attention would be significantly higher, which could lead to decisions on testing procedures being more political than professional.

National science testing should also help identify early students and schools in need of additional support.

According to the way we administer the test, it can be e.g. centralized and to be coded in one centre or to be left to teachers in schools. The latter is a common case where the national testing centre does not have sufficient capacity to collect and process all test materials. This choice, of course, reduces the control of the conditions in which the testing takes place and risks that the evaluation criteria are not uniform. It is certainly a better choice to have all the material with the answers delivered to the examination centre, which can then entrust the anonymized tests to a larger number of assessors / coders for further processing. Here are the critical logistical conditions, such as e.g. transport and storage of several thousand tests. For practical reasons, it is better to send only answer-sheets to the encoders, while a different procedure for postponing and storing tests is used for the tests themselves. It is even better if these answer sheets are digitized so that encoders can view them as documents on a computer or even online. Such a solution would require either someone to scan all the answer sheets or the whole test to be computerized. Due to the uneven conditions in different schools for taking tests on the computer, it is probably the colour of choice to continue the test in paper-pencil mode, and to digitize the answer sheets.

Due to the need to test the test as an instrument at the beginning, it is desirable that as many tasks as possible be open-ended to collect authentic student answers. Open answers require more effort around coding responses and training coders for this job. Ideally, teacher-coder training could be accredited in-service teacher training in the field of assessment, which would at the same time improve the practice of assessment in school and train coders, associates of the examination centre.

# Recommendations

To obtain the maximum information needed for further profiling of this study at an early stage of the national testing, it is necessary to gather as much useful information as possible through the test and accompanying questionnaires. Here are the most important recommendations for achieving this goal intended for the institution that prepares, conducts, and analyses testing (Examination Centre of Montenegro), the Ministry of Education and Bureau of Education Services.

* As there is already a national test for the subject of Mathematics, and for technique and technology we still do not have early versions of the national examination program at any level of education, the only sensible solution is to develop a national examination in the field of natural sciences or biology, physics, chemistry and (physical) geography. This testing would be best done in grade IX because that age is already covered by TIMS, PISA, and an external exam which provides additional opportunity to compare and develop procedures.
* The national science test program should include all subject outcomes and key competencies that are primarily developed through STEM disciplines, which can be checked with a paper-pencil test. It should be borne in mind that the test should not last longer than an hour, which means that the test cannot be more than 20-30 short tasks.
* For the national examination to have the necessary breadth, it is necessary for students to do a much larger number of tasks than one student can solve. It is therefore important that there are several different examination volumes that contain different tasks but are arranged in such a way that they overlap to a certain extent.
* A science test can cover all subjects in one exam notebook, but a better solution is to group the tasks relevant to certain subjects into clusters of tasks that are randomly arranged by exam notebooks. This means that some students, for example, do not have chemistry assignments in their exam notebook at all, while others do not have any geography assignments. What is important is that the sample for each of these clusters be representative.
* No matter what the tests look like at a later stage, it is important to start with as many open-ended tasks as possible to gather authentic answers and identify typical errors. For students, who are not too motivated, to answer an open-ended question, it should have as simple an account as possible. The best solution is "short answer" tasks.
* Bearing in mind that half of the students in Montenegro do not reach the second level of scientific literacy on the PISA test, the most important thing is that the national testing examines that level as much as possible in order to see what students really know or do not know. Most of the tasks on the national science test should be different variants of PISA tasks from the second level with a smaller number of tasks from other levels.
* Coding answers to open-ended tasks is not quite simple and therefore it is necessary to develop a series of examples for teacher-coder training through which teachers would be trained in coding and harmonize criteria.
* It is especially important that teachers play an active role in the national survey. Otherwise, they will perceive the whole examination process as an unnecessary obligation or as an attempt by the education authorities to control the results of teachers' work. They can have their role in all phases: from writing assignments and checking the characteristics of those assignments to coding the results, training their colleagues, interpreting the results, and writing recommendations for improving teaching.
* In order for the competency test to be more complete, it is necessary that the test is accompanied by at least two questionnaires: for students (different background variables, motivation, self-confidence, attitudes, plans for continuing education, etc.) and for teachers (e.g. teaching methods, curriculum, priorities and attitudes of teachers). As competencies include skills, attitudes, and values ​​in addition to knowledge, it is equally important for this examination what students think about STEM topics as well as what they know from it. International studies PISA and TIMSS have well-developed questionnaires of this type that need to be adapted for national testing purposes.
* Tasks from national testing should remain confidential and should not be shared with associates or made public. This does not apply to a small number of quality tasks (10-20%) that can be used for teacher training and public presentation of the test, or test results. The expert team that analyses the test results should select the tasks that should be published while keeping all others secret for the needs of the next test cycle. The practice of storing tasks and their characteristics is a condition for the results of tests in two consecutive cycles to be comparable. The tasks should be pre-tested, because in that way the established characteristics of the tasks (validity, discrimination and reliability) would confirm which task is of good quality and which is not, i.e. which of the tasks should be used in testing.
* Finally, in order to implement a quality and long-term testing programme, it is necessary to invest in the capacity of the Examination Centre, including human and material resources as well as facilities. Especially important are the automation and digitalization of the process, which would reduce the costs of national testing in the long run, as well as the human resources – the people who manage the process, analyze and competently interpret the data.

1. The test specification refers to the specific choice of tasks in a particular test and contains data such as the unique task number, task name, author name, task type, area to which it belongs, expected level of difficulty, etc. After processing the test results, these data are joined by empirically obtained statistical parameters. [↑](#footnote-ref-1)
2. The level scale is not the same for math and science. Therefore, the data "below level 1" for science is the sum of the share of students below level 1b and at level 1b on the scale for science. [↑](#footnote-ref-2)