NACIONALNI CENTAR ZA VANJSKO VREDNOVANJE OBRAZOVANJA

National Centre for External

Evaluation of Education

The Republic of Croatia

Development of Instruments in Croatian National Assessment

Project Report

Zagreb, November 2009

NATIONAL CENTRE FOR EXTERNAL EVALUATION OF EDUCATION

Development of Instruments in Croatian National Assessment

Project Report

PUBLISHER:

National Centre for External Evaluation of Education

FOR THE PUBLISHER:

Goran Sirovatka, B.Sc., director of *National Centre for External Evaluation of Education*

EDITORS:

Sanja Fulgosi, Ph.D.

Natalija Gjeri, B.A. History and Archaeology

AUTHORS:

The Centre staff: Sanja Fulgosi, Ph.D. Natalija Gjeri, B.A. History and Archaeology, Jasmina Muraja, Ph.D., Biljana Vranković, M.Sc., Natalija Ćurković, B.A. Psychology, Josip Šabić, B.A. Psychology

External experts: Maja Reberšak, B.A. Teacher, Željka Milin Šipuš, Ph.D., Ljubica Stanković, B.A., Miroslav Smuđ, B.A., Sanja Antoliš, B.A., Jelena Gusić, B.A., Dragica Martinović, B.A., Kristina Penzar, B.A., Josipa Pavlić, B.A., Jagoda Krajina, B.A., Dobrila Golubović, B.A.

EXPERT CONSULTANT (for chapters: 1, 2, 3, 4, 5, 7):

Zoran Žitnik, B.A. Psychology

LANGUAGE EDITOR:

Irena Ivančić, B.A. Croatian language

PROOFREADING:

Biljana Vranković, M.Sc.

GRAPHIC DESIGN:

Zoran Žitnik

PRINT:

Graphprint d.o.o., Čakovec

ISBN 978-953-7556-10-5

CIP record is available in the computer catalogue of the National and University Library in Zagreb under No. 719627.

We would like to thank all the school directors, the test coordinators, the other employees of the schools, and the students who have readily opened the doors to their schools and enabled a successful implementation of *National Assessment of Student Achievements* during the school year 2008/2009.

Your cooperation with the *National Centre for External Evaluation of Education* is extremely valuable for the realisation of projects that bring great changes in the sense of promoting quality in the educational system of the Republic of Croatia.

CONTENTS

INTRODUCTION

Jasmina Muraja, Natalija Gjeri

Developing a system of external evaluation of student achievements in education[1] is one of the strategic goals of the Republic of Croatia, as described in the document 'Education Sector Development Plan 2005-2010' published by the Ministry of Science, Education and Sports (2005). All of the priorities and aims listed in the *Plan* have arisen from the need to transform the Croatian education system, and have been harmonized with the standards in the European Union education system. The document highlights, with the purpose of improving the quality and efficiency of the education system, that: 'A system of external evaluation with standardized procedures will be established to develop, manage, and improve the monitoring of the achieved knowledge and skills. The evaluation will also cover the overall work of schools and teaching staff.' (Ministry of Science, Education and Sports, 2005; p.11). In accordance with that goal, *the National Centre for External Evaluation of Education*[2] (hereinafter *the Centre*) was founded, with the task of introducing external evaluation of student achievements in education into the Croatian education system in the form of national assessment.

National assessment is in the form of standardized external tests in one or more subjects conducted at the same time, in the same conditions, and with the same criteria for all students. Since 2006, the *Centre* has been conducting national assessment in gymnasium high schools, and since 2007, in vocational high schools. According to its goal and structure, the assessment has been preparing the students for the state Matura that is to be conducted in the school year 2009/2010.

For the purpose of improving the process of external evaluation, the *Centre* has recognized the need to develop and introduce contemporary models of standardized examinations that are being conducted worldwide. Since February 2007, designing the project related to the development of the system of external evaluation has begun, within the framework of the international Pre-Accession Projects Programme MATRA, financially supported by the *Ministry of Foreign Affairs of the Netherlands*. The project called *Development of Instruments in Croatian National Assessment*[3] (hereinafter *the Project*) was approved and realized in cooperation with the Dutch Institute for Educational Measurement CITO from April 2008 to September 2009. The lessons learned from this *Project* are the basis for the development of national assessment in the Republic of Croatia.

The *Project* coordinator and user is the *Centre*, and the *Project* has been realized with the support and partnership of the *Ministry of Science, Education, and Sports* (hereinafter the *MSES*).

This *Report* presents information on the aims, activities, and results of the *Project*, and is intended for the *Centre* staff and all of their external experts. It is directed at all responsible persons and institutions of the

[1] All of the terms used in this *Report* are masculine neutral and refer to persons of both male and female gender. In case of specific reference to the female gender, the terms will indicate so.

[2] It was founded in 2004, according to the Law on National Centre for External Evaluation of Education, Official Gazette, No. 151/2004.

[3] The original name of the project is 'Development of Instruments in Croatian National Assessment', Matra Pre-Accession Projects Programme (MPAP), NL

education system. The described course of the *Project* and the results presented will guide the experts from the *Centre* and the *MSES* in the planning of the further development of the system of external evaluation of student achievements. Finally, this document is important for all schools, their directors, teachers, and experts, in familiarizing them with the results of the implemented *Project* and with the development of standardized external evaluation tests used in the system.

REPORT OVERVIEW

This *Report* describes the mentioned *Project*, that is, the development and the implementation of the external evaluation tests, which forms the central part of this *Project*. The Report comprises eight chapters, which describe all of the *Project* phases.

The first chapter describes the general and the specific aims of this *Project* and the activities that implemented the aims and the results of the *Project*.

The second chapter discusses the selection of the subjects, the Croatian language and Maths, which were tested, and the subject domains selected for testing.

The third chapter describes in detail the assessment instruments created for the needs of this *Project*: the tests and the background questionnaires; specifies the information about the selection of schools and students and about the methods of administering the tests.

The forth chapter explains the general principles of score interpretation for the subject domains.

The fifth and the sixth chapters present how the subject domains in the Croatian language and Maths were assessed using the tests.

The seventh chapter describes the course of the development and the results of the background questionnaires.

The eighth chapter describes the psychometric data analysis conducted after the assessment in the subjects of the Croatian language and Maths. The psychometric analysis was done according to the principles of the *Classical Test Theory* (CTT) and the *Item Response Theory* (IRT).

1. THE PROJECT *Development of Instruments in Croatian National Assessment*

Jasmina Muraja, Natalija Gjeri

1.1. THE *PROJECT* AIMS

General aims of this *Project* are the following: improving the quality of the education system by introducing external evaluation in the form of national assessment, and harmonizing the Croatian education system with the European Union programmes.

Specific aims of this *Project* are the following: educating the *Centre* staff in external evaluation and educating the external experts (work groups) in the development of testing materials.

According to the *Project* aims, the *Centre* staff, in cooperation with the external experts and monitored by the experts from the Dutch Institute for Educational Measurement CITO, have passed all of the phases in the development, implementation and analysis of the national assessment. This assessment differs from the previous ones, which have been conducted so far as a form of preparation for the state Matura exam, in the aims, structure and the sample (*Table 1*). To indicate the difference, this assessment was named ***National Assessment of Student Achievements****. National Assessment of Student Achievements* was conducted in two subjects: the Croatian language and Maths. Apart from the tests, two background questionnaires were designed that the students completed after the assessment.

The aims of the *Project* also included educating the *Centre* staff in the area of *Classical Test Theory*, and training them to develop and analyse the tests according to the *Item Response Theory.*

The scope of the *Project* determined the assessment to be conducted on a sample of gymnasium programme students. It was estimated that the second grade is the time when the acquisition of the selected educational outcomes is completed, and therefore it was decided for the assessment to be conducted with second grade students.

Table 1. Differences between the national assessment conducted so far and the *National Assessment of Student Achievements*

	National assessment preparing for the state Matura	***National Assessment of Student Achievements***
Aims	- introducing into the system standardized tests that assess the student achievements at the end of high school education	-developing the methodology for the assessment of student achievements in specific subject domains from the subject curriculum
Subject domains	- three obligatory subjects (Croatian, Maths, foreign language) and elective subject(s) - the assessment can include the whole material planned in the Gymnasium Curriculum summarized in the Assessment Catalogue	- the assessment comprised specified subject domains within the curricula for the Croatian language and Maths - this assessment included also the subject domains not elaborated in the curricula
Sample	- all high school students took part in the national assessment	- *National Assessment of Student Achievements* was conducted with a sample of second grade gymnasium program students
Psychometric analysis	- according to the principles of the *Classical Test Theory*	- according to the principles of *Classical Test Theory* and *Item Response Theory*
Results	- students were informed on whether and up to what level they had achieved the competences necessary for graduating from high schools and continuing the education	- methodology for evaluating the level of student knowledge and competence developed - insight into the structure of the achievement of the specifically defined subject domains in the curricula

1.2. DESCRIPTION OF THE ACTIVITIES OF THE *PROJECT*

The Project was conducted **from February 2008 to September 2009** through a series of activities specified in *Tables 2* and *3.* The development phase of the *Project* started in February 2008 after the talks of the *Centre* staff and the representatives from the *MSES* with the experts from CITO. After the initial talks related to the *Project* aims, in April 2008 the initial conference '*Development of Instruments in Croatian National Assessment*' was held, where the *Project* was presented to the *Centre* staff, to the external experts and to the numerous representatives from the educational institutions. The experts from CITO held lectures on the importance of the *National Assessment of Student Achievements* in monitoring and improving the educational system.

Table 2. The course of the *Project* activities during the year 2008

Time Period	Activity
5.2. – 9.2.2008.	• CITO experts at the *Centre* • initial meetings of the *Project* managers and coordinators
15.4.2008.	• conference 'Development of Instruments in Croatian National Assessment' • presenting the *Project* and meeting the CITO experts
15.4. – 17.4.2008.	• forming groups that would take part in the *Project* • workshops in groups with the CITO experts ○ introducing the principles of the *National Assessment of Student Achievements*
18.4. – 1.6.2008.	• determining the aims and the structure of background questionnaires • designing questions for the background questionnaires
18.4. – 15.9.2008.	• test aims determined • selection of subject domains to be tested in the Croatian language and Maths • designing items for Croatian language and Maths
9.9. - 10.9.2008.	• cognitive laboratory: interviews with students in order to determine the reading ease and applicability of the questions from the background questionnaires

Time Period	Activity

15.9. - 18.9.2008.	• CITO experts in the *Centre*
	• workshops in groups with CITO experts:
	  ○ overview of the test structure, finishing the items, test design for pre-testing
	  ○ overview of the questions and the structure of the background questionnaires for pre-testing

19.9. - 26.9.2008.	• selection of schools and students for pre-testing
22.9. - 17.10.2008.	• preparing test booklets and background questionnaires for graphic design; proofreading, graphic design of the test booklets and background questionnaires, corrections
20.10. - 31.10.2008.	• printing the test booklets and background questionnaires • preparations for pre-testing: informing the schools, meetings with the county test coordinators, preparing the instructions for pre-testing
3.11. - 15.11.2008.	**PRE-TESTING**
17.11. – 24.11.2008.	• marking and test booklet data input
24.11. – 27.11.2008.	• CITO experts at the *Centre* (review of the pre-testing) • workshop with the psychometrics group • preparation of data for the psychometric analysis of items
27.11. – 20.12.2008.	• psychometric item (test booklets) analysis according to the *Classical Test Theory* • input of background questionnaires results • analysis of background questionnaires
22.12. – 9.1.2009.	• presenting the results of pre-testing in the Croatian language and Maths according to the *Classical Test Theory* • analysis of the results • presenting the results of the background questionnaires analysis

Table 3. The course of the *Project* activities during the year 2009

Time Period	Activity

12.1. – 15.1.2009.	• CITO experts at the *Centre*
	• workshops in groups with CITO experts:
	  ○ test analysis according to the *Item Response Theory*
	  ○ the first selection of items for the main testing in the Croatian language and Maths
	  ○ suggestions of the tests designs in the Croatian language and Maths

15.1. – 30.1.2009.	• test design for the main testing • selection of schools for the main testing • informing the selected schools of the *National Assessment of Student Achievements*
2.2. – 25.2.2009.	• preparing test booklets and background questionnaires for graphic design; proofreading, graphic design of the test booklets and background questionnaires, corrections • selection of students in the schools in the sample • meeting with the test coordinators
23.2. – 6.3.2009.	• printing test booklets and background questionnaires
9.3.2009.	• test booklets and background questionnaires sent to schools
16.3. – 20.3.2009.	**MAIN TESTING ADMINISTRATION** **National Assessment of Student Achievements**
23.3. – 2.4.2009.	• marking and input of the test booklets results • input of background questionnaires results
3.4. – 20.4.2009.	• psychometric analysis according to the *Classical Test Theory* and the *Item Response Theory* • analysis of the background questionnaires results
9.4.2009.	• meeting of the *Project* supervisory board at the *Centre* (PAC meeting)
20.4. – 23.4.2009.	• CITO experts at the *Centre* • workshops in groups with the CITO experts: • analysis of the results obtained in the *National Assessment of Student Achievement*
24.4. – 1.9.2009.	• analysis of test and background questionnaires results ○ preparation of the *Project* report

1.9. – 21.9.2009.	• organizing the final conference
21.9. – 23.9.2009.	• CITO experts at the *Centre* • final conference: presenting the *Project* results
	***PROJECT* END**

At the beginning of the *Project,* four groups were formed, which developed the assessment instruments in the content and the structure: Croatian Language group, Maths group, Background Questionnaires group, and the Psychometrics group. *(Table 4).* The group for the Croatian language consisted of the *Centre* staff and experts in the subjects assessed (high school teachers and a university professor). The background questionnaires group and the psychometrics group consisted of the *Centre* staff.

Table 4. Groups that developed the assessment instruments within the framework of the *Project*

The Croatian Language group	The *Centre* staff: dr.sc. Sanja Fulgosi (coordinator) and Natalija Gjeri The Croatian language experts: Marina Čubrić, B.A. (Nadbiskupska klasična Gimnazija, Zagreb), Lidija Farkaš, B.A. (V. gimnazija, Zagreb), Višnja Jukić, B.A. (Ženska opća gimnazija Družbe sestara milosrdnica s pravom javnosti, Zagreb), Majda Bekić Vejzović, B.A. (V. gimnazija, Zagreb), Božica Jelaković, B.A. (XV. Gimnazija, Zagreb), Jugana Dagelić, B.A. (IV. gimnazija „Marko Marulić“, Split), B.A. dr. sc. Zrinka Jelaska (Filozofski fakultet Sveučilišta u Zagrebu, Odsjek za kroatistiku), Mirela Barbaroša Šikić, B.A. (Agencija za odgoj i obrazovanje, Zagreb)
The Maths group	The Maths experts: B.A. dr. sc. Željka Milin Šipuš (Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu), group leader, Josipa Pavlić, B.A. (Gimnazija Lucijana Vranjanina, Zagreb), Jagoda Krajina, B.A. (Tehnička škola Ruđera Boškovića, Zagreb), Dragica Martinović, B.A. (Ženska opća gimnazija Družbe sestara milosrdnica s pravom javnosti, Zagreb), Jelena Gusić, B.A. (XV. gimnazija, Zagreb), Sanja Antoliš, B.A. (XV. Gimnazija, Zagreb), Kristina Penzar, B.A. (Nadbiskupska klasična gimnazija, Zagreb), Dobrila Golubović, B.A. (Tehnička škola Ruđera Boškovića, Zagreb), Miroslav Smuđ, B.A. (Elektrostrojarska škola, Varaždin), Ljubica Stanković, B.A. (Osnovna škola Ksavera Šandora Đalskog, Donja Zelina)
The Background Questionnaires Group	*Centre* staff: Maja Reberšak (coordinator), mr. sc. Biljana Vranković, Kristina Svalina and dr. sc. Jasmina Muraja
The Psychometrics Group	*Centre* staff: Natalija Ćurković (coordinator) and Josip Šabić

After the initial Conference, the groups started the education through workshops led by the experts from the Institute for Educational Measurement CITO. Erna Gille educated and counselled the Croatian language group throughout the *Project*, Paul Van der Molen the Maths group, Johanna Kordes the background questionnaires group, and Frans Kleintjes the psychometrics group. At the very beginning of the *Project*, expert consultant Marius J. Ouborg counselled the *Centre* staff on the issues of the development of the information system. The Dutch coordinator was Jose Noijnos, and the leader of the *Project* was Erna Gille.

During the *Project*, there were two meetings of the Project Advisory Committee (PAC): Caroline van den Ende and Christina van der Heden from the Dutch organization EVD, Zrinka Čornij, representative of the Dutch embassy in Zagreb, Goran Sirovatka, director of the *Centre*, Jose Nojinos, CITO Project coordinator, Mihela Dubravac Šigir from MSES, later replaced by Monika Vričko from the same Ministry, Ana Šimunić and Stanka Crvik Orešković from the Central Office for the Coordination of International Funds, and Jasmina Muraja, PhD., the *Centre* Project leader. The Project Advisory Committee monitored

the implementation of the *Project* activities and evaluated positively its course and development. It influenced the decision of the Central Office for the Coordination of International Funds to highlight this *Project* as an example of good practice in the year 2008, so the *Project* was shortly presented at the annual presentation of the new round of the Dutch bilateral programmes in the year 2009.

1.3 THE *PROJECT* RESULTS

The *Project* resulted in the following:

- the *Centre* staff was educated with the help of the experts from the Dutch Institute for Educational Measurement CITO and they improved their knowledge in the area of item writing and the development of assessment instruments – the tests

- the *Centre* staff was educated with the help of the experts from the Dutch Institute for Educational Measurement CITO and they improved their knowledge in the area of constructing the background questionnaires questions and the construction of the questionnaire itself

- the *Centre* staff was educated with the help of the experts from the Dutch Institute for Educational Measurement CITO and they advanced their knowledge in the area of psychometrics - the development of assessment instruments (tests and background questionnaires) and the analysis of the instruments according to the *Classical Test Theory* and the *Item Response Theory* (IRT)

- experts in the subjects of the Croatian language and Maths were educated with the help of the experts from the Dutch Institute for Educational Measurement CITO and they improved their knowledge in the area of item writing and test construction

- **for the first time in Croatia, formative assessment was conducted: *National Assessment of Student Achievements***

- the results of the national assessment in Croatian language and Maths were analyzed according to the *Item Response Theory* for the first time in Croatia

- the basis for the development of future national assessment in Croatia has been set, with the aim of promoting the quality of education.

2. SELECTION OF SUBJECT DOMAINS FOR THE *NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENTS*

Sanja Fulgosi, Željka Milin Šipuš

2.1. SELECTION OF SUBJECTS TO BE ASSESSED

Sanja Fulgosi

The Republic of Croatia, as it was highlighted in the introduction, is leaning toward a high-quality, efficient educational system that will enable an individual, through lifelong learning, to gain better prerequisites for employment, for personal education and for the betterment of the society.

An important role in the construction of such a high-quality educational system is played by the system of external evaluation and assessment that can, through reliable factors, enable an objective insight into the efficiency of individual components of the educational system, as it was highlighted in the mentioned document *Education Sector Development Plan 2005-2010* (MSES, 2005). The project *Development of Instruments in Croatian National Assessment*, in whose framework the national assessment of student achievements was conducted, contributes to the systematic development of the process of external evaluation of educational achievements in the Republic of Croatia. The strategy of the national examinations should be developed simultaneously with the application of the contemporary theoretical approaches in test design, and with a well-planned selection of subject domains.

Ever since the meeting in Lisbon in 2000, the European Council has been intensively defining the guidelines, the objectives, and the standards for a successful lifelong learning as the main prerequisite for a high-quality modern education during this decade.

By analysing the objectives of a systematic education, it has been concluded that a prerequisite for acquiring knowledge is developing elementary skills, the basic ones being literacy and numeracy. After that, the notion of elementary skills was considered and fitted into the wider framework of key competences in the document *Education and Training 2010 Work Programme* issued by the European *Commission Directorate General for Education and Culture* (2004). Due to the key points of its content, the document can be considered *a framework for achieving key competences.*

The eight competences specified in the framework are the following: communication in mother tongue, communication in a foreign language, mathematical literacy and basic competences in science and technology, digital competence, learning to learn, social and civic competences, sense of initiative and entrepreneurship, cultural awareness and expression.

When it comes to evaluation and achievement assessment, it is important to note that the measurement and the evaluation of competences can be a very complex task, which presupposes the application of special instruments. This *Project* therefore does not aim to evaluate competences but the level of knowledge, which is an important premise for achieving competences.

The Project *Development of Instruments in Croatian National Assessment,* is a step towards a systematic national achievement assessment, the results of which can be one of the indicators whether the Croatian educational system is following contemporary requirements in education. That is why the *Project* starts with the subjects of the Croatian language and Maths, as they provide for the basic skills of literacy and numeracy, that is, they lead towards a minimum of two out of eight key competences: *communication in mother tongue*, and *mathematical literacy and basic competences in science and technology*. The starting point of these competences is in the subjects of the Croatian language and Maths. However, at the same time, it needs to be noted that a characteristic of all key competences is their quality of being *generic* and

transversal, which means that they, to a large degree, correlate with all subjects, and are a prerequisite for acquiring all other competences. That very fact emphasizes the Croatian language and Maths as basic subjects whose content is systematically acquired from the beginning of education.

All of the mentioned facts emphasise the importance of Croatian and Maths and demonstrate that these two subjects make a good content starting point in the systematic design and development of national assessment.

2.2. CROATIAN – SELECTION OF SUBJECT DOMAINS

Sanja Fulgosi

One of the eight key competences is *communication in mother tongue,* which is defined as the 'ability to express and interpret thoughts, feelings and facts in oral and written form (listening, speaking, reading and writing) and the appropriate linguistic interaction in a range of social and cultural topics – education, work, home and free time' (European Commission Directorate General for Education and Culture, 2004). The knowledge, skills and attitude that this competence implies in the mother tongue, are listed in the already mentioned document that brings forth a *framework of key competences.* Some of the knowledge and the skills from this European document coincide with the objectives, purpose and content indicated in the *Curricula for Gymnasium High Schools, Croatian for Gymnasium High Schools* (Ministry of Culture and Education, 1994) which was the starting point for assessment in the Croatian language. Within the framework of key competences it is highlighted that the important thing in the mother tongue is a 'solid knowledge in basic vocabulary, functional grammar, and style and language functions', and equally important being the 'awareness of the different types of literary texts and their characteristics, and of the types of non-literary texts and their characteristics' (European Commission Directorate General for Education and Culture, 2004).

Four subject domains within the subject of Croatian language have been selected: grammar, *theory of literature*, and reading literary and non-literary texts.

Some areas in *grammar* and *theory of literature* are studied as early as lower grades elementary school, but they are systematically organized in the gymnasium curriculum so the subject domains followed Croatian for gymnasium high schools in *Curricula for Gymnasium High Schools* for the first and the second grade. The area of grammar, that is, familiarity with the linguistic system and functional language rules, is important for communication in the Croatian language, and the terms from the *theory of literature* are the basis for understanding *a literary text.*

Acquiring and developing competences in all four language skills (reading, writing, listening and speaking) is a general objective of teaching Croatian. It is of importance to assess achievements in all four language skills, and this *Project* developed the procedure for assessing the reading skill in two subject domains*: reading a literary text and reading a non-literary text.*

A large part of the Curriculum, and of teaching Croatian, is dedicated to developing the competence of reading a literary text, and was therefore important to include the subject domain *reading a literary text* as a separate area of assessment in the subject of Croatian language.

One of the objectives of testing the subject matter in Croatian was to assess the acquisition of that subject matter that was not elaborated in the *Curricula for gymnasium high schools, Croatian for high schools*. The mentioned *Framework for Key Competences* emphasises the importance of communication in different social, cultural, business, or personal situations. That is where that document overlaps with the general objectives and the purpose indicated in the *Curricula for gymnasium high schools, Croatian for high schools,* emphasising the importance of communication in the Croatian language in all functional styles.

In explaining the purpose that the Croatian language serves as a school subject, one element was highlighted, which corroborated the selection of the subject domain *reading a non-literary text* in this testing. It states that it is important to 'acquire language knowledge, language culture and competences to use the Croatian language in all text types, functional styles and means of communication' (Ministry of Culture and Education, 1994, p.151).

Based on everything mentioned above, four subject domains were selected for this testing.

Table 5 outlines all of the subject domains and the number of items referring to a particular domain.

Table 5. Subject domains and the number of items in Croatian

Subject Domain	Number of Items
grammar	15
theory of literature	15
reading a literary text	15
reading a non-literary text	20
TOTAL	**65**

2.3. MATHS – SELECTION OF SUBJECT DOMAINS

Željka Milin Šipuš

National Assessment of Student Achievements in the subject of Maths was conducted on the basis of *Curricula for gymnasium high schools*, chapter *Maths* (Ministry of Culture and Education, 1994).

In the school year 2006/2007, the *Curricula for gymnasium high schools* included the guidelines from the Croatian National Education Standard (Ministry of Science, Education, and Sport, 2006). The second grade gymnasium program students who took part in the *National Assessment of Student Achievement*s experienced these changes in the curriculum when they were in the eighth grade of elementary school, and this fact was taken into consideration in the selection of subject domains. In order to select subject domains, the Maths subject matter that is studied according to the curriculum before the second year in high school, was divided into four Maths domains and into the correspondent sub-domains. That resulted in the possibility of dividing the Maths subject matter that is studied before the second year in high school, into the following thirteen sub-domains:

a) mathematics basics

1. percent and ratio
2. mathematical literacy

b) algebra

3. fractions
4. powers and roots
5. factoring and brackets
6. complex numbers

c) functions and equations

7. linear equation
8. linear function
9. quadratic equation
10. quadratic function
11. systems

d) geometry

12. triangle
13. areas (measuring).

The selection of subject domains for the *National Assessment of Student Achievements* was based on the presence of a particular domain in the *Curricula for gymnasium high schools, Maths,* on the interest for the domain, and special importance was given to the domains studied in the second grade in high school. Taking into consideration everything mentioned above, six subject domains were chosen to test Maths: fractions, linear equation, quadratic equation, quadratic function, triangle, and mathematical literacy.

For other subject domains, for instance, *areas,* it was decided that they should be tested while included in the selected subject domains (*triangle)*, and more complex geometrical matters, as well as matters of percentages and ratios, within the subject domain of *mathematical literacy.*

The subject domain of *mathematical literacy* was highlighted as a special domain with the purpose of establishing the competence of applying mathematical ideas in real-life situations. The development of that competence is one of the key objectives of education in Maths.

Mathematical literacy is one of the key competences of lifelong learning according to the document *'Implementation of the work programme 'Education 2010'* of the *European Commission Directorate-General for Education and Culture* (2004). The document defined mathematical literacy as the 'competence in adding, subtracting, multiplication, division, and using ration in reflexive and written calculating in solving different everyday situations problems' (the European Commission Directorate-General for Education and Culture, 2004). It is implied that there is previous knowledge of numbers, measures, basic calculation methods, and of using elementary forms of mathematical representation (graphs, formulas, statistics) in everyday situations. Moreover, it is implied that there is previous knowledge of mathematical terms and concepts, including the most important theorems in geometry and algebra.

Mathematical literacy is also one of the three subject domains of the international testing PISA. For the purpose of that testing, mathematical literacy is defined as the 'student's competence to efficiently analyse, discriminate and communicate ideas in forming, formulating, solving and interpreting mathematical problems in different situations' (Braš Roth, Gregurović, Markočić Dekanić, Markuš, 2007, p. 122).

According to the Maths test design, items from the subject domain of *mathematical literacy* were represented in all test booklets. In the *Curricula for gymnasium high schools, Maths,* these items appear as application items, that is, textual items (Ministry of Culture and Education, 1994). For the purpose of a clearer assessment of understanding individual mathematical concepts and skills, none of the subject domains, apart from *mathematical literacy,* contained items that would link the subject matter from different subject domains.

The Maths work group conducted a preliminary testing in order to establish the number of items that students can solve in the given time. Ninety-nine students from several high schools were involved in the test and it was determined that between one and two minutes were needed per given item. This time period was taken into consideration for individual types of items in the final selection of item clusters and the test design.

Within each subject domain the so-called *study directions* were established – clusters of certain items testing the same mathematical concept or skill, and within these clusters it was possible to order the items by level of difficulty.

For each of the subject domains, around a hundred items were written, and the items for the main testing were chosen from among them, according to the psychometric parameters. The number of items in individual subject domains is presented in *Table 6.*

Table 6. Subject domains and the number of items in Maths

Subject Domain	Number of Items

fractions	24
linear equation	24
quadratic equation	24
quadratic function	27
triangle	27
mathematical literacy	24
TOTAL	**150**

3. THE STRUCTURE AND THE IMPLEMENTATION OF NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENTS

Natalija Gjeri

3.1. ASSESSMENT INSTRUMENTS

As it was stated in the first chapter, the aims of the project *Development of Instruments in Croatian National Assessment* is to educate the *Centre* staff in the area of external evaluation of the students' educational achievements, and to educate the external experts, experts in the subjects that were tested, Croatian and Maths, in the area of item writing. Within the framework of the education **tests in Croatian and Maths** were created as well as **Background Questionnaires** that the students completed with each of the test booklets. The items and the questions in the Background Questionnaires were pre-tested in the third year of high school, and the main testing (National Assessment of Student Achievements) in the second year of gymnasium program high schools.

3.1.1. Work Groups for Item Writing and Assessment Instruments Construction

According to the objectives of the *Project*, four work groups were formed:

a) Croatian work group

b) Maths work group

c) Background Questionnaires work group

d) Psychometrics work group.

The Croatian and Maths work groups' tasks were the following: selecting subject domains, determining educational outcomes, and writing the items for Croatian and Maths. The Background Questionnaires work group's tasks were: determining the aims and the structure of the background questionnaires and writing the questions, in cooperation with the Croatian and Maths work groups. The Psychometrics works group's tasks were: test design in cooperation with the Croatian and Maths work groups, analysing the tests and the background questionnaires. All four groups cooperated in order to have high-quality, reliable and valid assessment instruments.

All of the four work groups were counselled by experts from the Dutch Institute for Educational Measurement CITO, who were in charge of educating and guiding the groups.

3.1.2. Tests and background questionnaires

The instruments developed for the purpose if this *Project* are the following:

a) tests in the subjects of Croatian and Maths

The tests were meant to verify the knowledge and skills of the **second grade gymnasium program students** in narrowly defined domains of the *Curricula for Gymnasium High Schools* for Croatian and Maths (ministry of Culture and Education, 1994). In addition, the tests were meant to check the domains not elaborated in the Curricula, particularly the domain of reading a non-literary text in Croatian and the domain of mathematical literacy in Maths.

b) background questionnaires

Along with the tests, two background questionnaires were constructed:

1. the background questionnaire to be completed after the Croatian test

2. the background questionnaire to be completed after the Maths test

Both questionnaires had a common part, the first general part, which checked the gender, age, and socio-economic status of the students in the sample. The second part of the questionnaire aimed at the attitudes and opinions of the students related to the subject tested, Croatian or Maths.

3.1.3. Pre-testing of test items and background questionnaire questions

The pre-testing of items for the tests in Croatian and Maths, and of the background questionnaire questions was conducted in **November 2008** with a sample of **third grade** students from 33 gymnasium high schools (1186 students did the Croatian test, and 1229 the Maths test). The psychometrics group analysed the items according to the principles of *Classical Test Theory* and *Item Response Theory*. After a detailed analysis of the data received, the Croatian and Maths work groups selected the psychometrically valid items for the main testing.

The criteria for the selection of items for the main testing were the following:

- content validity of the items – those items that best represented the subject domains were selected
- psychometric characteristics – items that were selected were those that showed a satisfactory index of discrimination (items discriminating students with different abilities) and that satisfied a certain difficulty level
- the distribution of items according to the level of difficulty – in each subject domain and each test booklet, items of different difficulty levels were selected, in five categories, from very easy to very difficult, with most of the items being at a medium difficulty level
- the satisfactory number of items that represented the subject domain – a minimum of 15 items were selected for each subject domain

Along with the test items, the background questionnaire questions were pre-tested. The analyses conducted resulted in the measurable characteristics of the questions and the factor structure of the questionnaire. After a detailed analysis, the background questionnaires work group, in cooperation with the Croatian and Maths work groups, selected and grouped the questions for the main testing.

3.1.4. The main testing : *National Assessment of Student Achievements*

National Assessment of Student Achievements was conducted in the **second year of gymnasium high schools** in the subjects of **Croatian and Maths.** With each of the test booklets, the students completed the corresponding background questionnaire.

A. The Croatian language test

The test design for the Croatian language consisted of **three test booklets,** and they contained:

a) 4 subject domains

b) 12 units (clusters)

c) 65 items. [4]

The four subject domains tested were the following: *theory of literature*, reading a literary text, reading a non-literary text, and grammar. The prompts for the first three subject domains were texts: for *theory of literature* and reading a literary text there were literary texts, and for reading a non-literary text, there was a non-literary text. Each text dealt with one unit (a cluster), and the subject domain of grammar was a cluster without a single starting text.

Test structure

The test was composed of items that were organized into smaller units, that is, clusters that belong to a particular subject domain. Each unit was tested in two test booklets, and the booklets varied in the number of items: the first test booklet had 40 items, the second one 48 items, and the third one, 42 items. In two test booklets, all of the subject domains were covered, and in one test booklet all of the subject domains except grammar. Regarding the test design, each of the items was repeated in two test booklets. For each of the three booklets, the set time for completing the test was 60 minutes.

The items were both closed and open type. The closed type items were multiple choice questions, where students choose one of the three or four answers suggested. The open type items were short answer items and completion items, where students offer one word or a group of words as the answer, and the extended reply type where the answer consisted of one or two sentences.

B. The Maths test

The Maths test design consisted of **seven test booklets**, and they contained:

a) 6 subject domains

b) 22 units (clusters)

c) 150 items [5]

The six subject domains tested were the following: quadratic equation, triangle, fractions, linear equation, quadratic function, and mathematical literacy.

Test structure

The test was composed of items that were organized into smaller units, that is, clusters, which were organised into test booklets. Each unit had six or nine items, and each of the seven booklets had 48 items. Not all of the test booklets contained items from all of the subject domains; three of them contained items from the domains of quadratic equation, triangle, and mathematical literacy, while the other four contained items from the subject domains of fractions, linear equation, quadratic function, and mathematical literacy.

[4] In the fifth chapter *Croatian Language – subject domains* – only the items fitting the test model were described

[5] In the sixth chapter *Maths – the subject domains* – only the items fitting the test model were described

Regarding the test design, each of the items was repeated in two test booklets, and the items belonging to the domain of mathematical literacy in three or four test booklets. For each of the three booklets the set time for completing the test was 60 minutes.

The items were both closed and open type. The closed type items were multiple choice questions where students choose one of the four answers suggested, and the complex multiple choice questions where students choose more than one of the four answers suggested. The open type items required the student to offer a solution to a particular item. The students were allowed to use a pocket calculator.

C. Background Questionnaires

1. The Background Questionnaire that was filled after the Croatian test was composed of two parts:

a) The general part, which surveyed the gender, age and socio-economic status of the students, and was composed of 17 questions. That first, general part in the background questionnaire is the same as the one that was completed after the Maths test.

b) The part related to the subject of Croatian language consisted of 22 questions. The questions referred to the motivation of students to study the Croatian language, the studying and the teaching of the Croatian language, the book reading and the applying of the things learned in everyday life.

2. The Background Questionnaire that was completed after the Croatian test was composed of two parts:

a) The general part, which surveyed the gender, age, and socio-economic status of the students, was composed of 17 questions. That first, general part in the background questionnaire is the same as the one that was completed after the Croatian test.

b) The part related to the subject of Maths consisted of 22 questions. The questions referred to the motivation of students to study Maths, the extra-curriculum students' work, the teacher's work, the teaching methods and techniques, teaching aids used in classes, and the application of the things learned in everyday life.

3.2. SELECTION OF SCHOOLS AND STUDENTS FOR THE *NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENTS*

3.2.1 Population and sample type

The scope of the *Project* required the *National Assessment of Student Achievements* to test the knowledge and skills of **second grade gymnasium program students.**

In the *National Assessment Of Student Achievements* a **two-stage implicit stratified sample** was used. In the first stage, 44 schools were selected, and after the selection of the schools, the students in those schools were selected. Taking into consideration the fact that the Maths test design consisted of more test booklets than the Croatian test design, the number of schools to be involved in the Maths test is higher than the than the number of schools involved in the Croatian test. The Croatian test was conducted in 40 schools, and the Maths test in 44 schools. According to that, the Maths test was taken by more students than the Croatian test; by 1244 of them, while 990 students took the Croatian test.

3.2.2. School selection

In selecting the schools, the following stratification variables were used:

1) Gymnasium program high school (gymnasiums and mixed schools). Two types of schools were included in the assessment: gymnasiums and mixed schools. Gymnasiums are high schools to which the *Curricula for High Schools* is applied, and mixed schools are those that have, alongside the gymnasium curriculum, also the *Curriculum for Vocational Schools.* In mixed schools, only the gymnasium program students were selected for the assessment.

2) Division according to regions (six regions). The counties of the Republic of Croatia are grouped in six regions according to their geographic position.

3) Level of urbanization (city, town, place). The places in Croatia are divided according to the number of population into the categories of cities, towns, and places.

Tables 7, 8, and 9, show the distribution of schools according to the mentioned stratification variables.

Table 7. Distribution according to the gymnasium curriculum school type

School Type	The number of schools in the sample (Croatian)	The number of schools in the sample (Maths)
gymnasium	23	27
mixed schools with gymnasium curriculum	17	17
TOTAL	**40**	**44**

Table 8. Distribution of schools according to regions

Regions	**Counties**	**The number of schools in the sample (Croatian)**	**The number of schools in the sample (Maths)**
central Croatia	Zagrebačka, Karlovačka, Sisačkomoslavačka, Virovitičko-podravska, Koprivničko-križevačka, Bjelovarskobilogorska	8	10
east Croatia	Vukovarsko-srijemska, Osječko-baranjska, Požeško-slavonska, Brodsko-posavska	6	7
north Croatia	Krapinsko-zagorska, Međimurska, Varaždinska	3	3
west Croatia	Primorsko-goranska, Ličko-senjska, Istarska	6	6
south Croatia	Šibensko-kninska, Splitsko-dalmatinska, Zadarska, Dubrovačko-neretvanska	10	11
the city of Zagreb	the city of Zagreb	7	7
TOTAL	**40**	**40**	**44**

Table 9. Distribution of schools according to the level of urbanization

Level of urbanisation	The number of schools in the sample (Croatian)	The number of schools in the sample (Maths)
city	12	13

town	16	19
place	12	12
TOTAL	**40**	**44**

3.2.3. Selection of students

The schools that were included in the sample sent the *Centre* data about the second grade students. After that, the selection of students in the schools followed. In the schools that had more than 70 students (gymnasiums), 35 students were selected systematically to do the Croatian test and 35 students to do the Maths test. In schools that had less than 70 students in second grade (mixed schools), all of the students were included in the sample. Half of them were selected to do the Croatian test, and half of them were selected for the Maths test. In four additional high schools, there was no Croatian test, but all of the 70 students did the Maths test. In selecting the students, **the gender** variable was taken into consideration.

1021 students were selected to do the Croatian test (57% female, and 43% male students), and **1295 students** were selected to do the Maths test (56% female students, and 44% male students).

3.3. *NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENTS*

National assessment of student achievements was conducted **between 16th March 2009 and 20th March 2009**. The following activities preceded the assessment:

- informing the directors and test coordinators in 44 schools that participated in the assessment
- meetings with the test coordinators in 44 schools that participated in the assessment
- coding the test booklets
- seating the students in the test rooms according to the test booklets
- printing the materials and sending them to schools.

The test booklets and the background questionnaires were assigned to the schools based on a system of equal distribution. Each of the 40 schools was given each of the three test booklets in Croatian and each of the seven booklets in Maths. Four schools, in which there was only the Maths assessment, also received each of the seven test booklets. The number of copies of one test booklet in a school depended on the number of students in that school who participated in the assessment. In the testing rooms, the test booklets were handed according to the pattern: Croatian 1-3, and Maths 1-7. The test booklets, as well as the correspondent background questionnaire, were coded, and the codes were assigned to schools and to each student that was involved in testing.

3.3.1. The role of the test coordinators

The test coordinators are persons appointed by the schools directors. They were, along with the other members of the *School Testing Committee,* in charge of conducting the assessment in the schools[6]. The participation of the test coordinators in the schools with a gymnasium curriculum in conducting the *National Assessment of Student Achievements* included the following:

- sending *the centre* data about second grade students so that the student selection could be done
- informing the selected students about the importance of the *National Assessment of Student Achievements*
- informing the students about the assessment procedure
- receiving, protecting, and storing the test materials
- supervising the assessment process and ensuring the regulation of the assessment procedure
- returning the test materials to *the Centre*

The test coordinators were given instructions for conducting the assessment, the form for conducting the assessment, the directions and the form for the test administrators that they appointed. In order to meet the needs of the schools, the *Centre* staff arranged the assessment dates and times with the test coordinators.

[6] The role of the test coordinators was regulated in the *State Matura Regulations*, Official Gazette, broj 87/08.

3.3.2. Duration of the Assessment

The assessment had different duration times for Croatian and for Maths.

1. Croatian: test duration 60 minutes, background questionnaire completion 10 minutes
2. Maths: test duration 90 minutes, background questionnaire completion 10 minutes

3.3.3. The rules of assessment

For the purpose of conducting the *National Assessment of Student Achievements,* the *State Matura Regulations* applies (Ministry of Science, Education, and Sports, 2008). The *Regulations* points referring to the process of conducting the assessment in the schools were adjusted for the needs of this particular assessment. The assessment was conducted according to the following points:

- The test coordinators were given the students' seating arrangements in the test rooms a week before the assessment. The students were to sit alone at a desk so that there were 15-18 students per room (the only exception being the larger halls).
- The test coordinator appointed the test administrators, a minimum of three days before the beginning of assessment.
- The seating schedule of students and test administrators in the testing rooms was made known by the test coordinator on the day of testing, 60 minutes prior to the beginning of the test, at a place easily reachable for the students. The seating schedule had to be put up at the entrance to the testing room.
- All of the teaching aids that could help the students needed to be removed from the test rooms.
- The students had to be present at the entrance to the test room a minimum of 15 minutes before the beginning of the test.
- The students were not allowed to bring mobile phones or other portable electronic devices to the testing room.
- During the test two test administrators were present, who could not be teachers in the subjects assessed. One of them was the test room manager.
- The test administrator had to respect all the regulations and instructions during the test.
- The test administrator noted down in the log all of the details that happened during the test.
- Considering the fact that the students had to complete the background questionnaire after writing the test, it was not allowed to exit the room before the time for the testing expired. After the test, the test administrator read additional instructions for completing the background questionnaires.
- The test administrator collected all the materials at the end of the writing and checked whether the students returned all of the materials and whether everything was done according to the instructions.

The *Centre* staff monitored the testing in most of the schools involved. After the end of the testing, the test coordinators sent the boxes with the test materials to the *Centre*, the forms they filled and the forms that

the test administrators filled. When the boxes with the test materials arrived to the *Centre*, the test materials were sorted according to the subjects, so that the scoring process could start.

3.4. THE STUDENT RESPONSE

990 students were tested in Croatian, and **1244 students** were tested in Maths. 31 students did not present themselves at the Croatian test, and 51 students did not present themselves at the Maths test. *Tables 10 and 11* show the main characteristics of the Croatian and Maths test samples.

Table 10. The number and the percentage of the students involved in the testing according to the school type

School Type	The number of students (Croatian test)	Percentage (Croatian test)	The number of students (Maths test)	Percentage (Maths test)
gymnasium	745	75	1004	81
mixed schools with gymnasium curriculum	245	25	240	19
TOTAL	**990**	**100**	**1244**	**100**

Table 11. The number and the percentage of the students involved in the testing according to the gender

Gender	The number of students (Croatian test)	Percentage (Croatian test)	The number of students (Maths test)	Percentage (Maths test)
Male	424	43	551	44
Female	566	57	693	56
TOTAL	**990**	**100**	**1244**	**100**

3.5. SCORING

The Croatian test booklets were scored by the Croatian work group, and the Maths test booklets by the Maths work group. Prior to the scoring, the groups, in cooperation with the Psychometrics group, defined the coding method and item scoring. For scoring the Croatian test, there was the *Guide for scoring open items* for each booklet, and it had clear criteria for scoring each answer. After the scoring, the results of the tests and the background questionnaires were psychometrically analyzed.

3.6. QUALITY ASSURANCE PROCEDURES

During the *Project*, while developing the *National Assessment of Student Achievements,* different procedures for ensuring quality were followed:

Education and monitoring by the external experts. Work groups that participated in the development of assessment instruments were educated and monitored by the experts from the Dutch Institute for Educational Measurement CITO.

Selection of work groups for item writing. The people who were involved in item writing are experts in the subjects assessed (Croatian and Maths). These experts were educated in item writing, and were lead by the *Centre* staff in order to ensure the quality and the validity of items.

Cognitive lab. Interviews with students were conducted in order to determine the legibility and the applicability of the background questionnaire questions, prior to the testing.

Pre-testing. The items and the background questionnaire questions were pre-tested in order to determine their validity for the main testing. The items for the main testing were selected according to the set criteria. It is also important to note that the Maths work group conducted an independent survey before the pre-testing in order to determine the number of items that the students can solve in the given time period. There were 99 students included in several schools, and the results obtained were taken into consideration in the final writing of units (clusters) and the test booklets for the pre-testing.

Standardized procedures of conducting the assessment. The assessment was conducted according to the set rules prescribed in the *State Matura Regulations*, with some smaller adjustments needed for this particular assessment.

Scoring. The scoring of closed and open question types in Croatian and Maths was conducted according to the set criteria. Taking into consideration the fact that there was a large number of open type questions in the Croatian test, also the *Guide For Marking Open Type Questions* was made for each of the test booklets. This was especially important for the extended answers type where it was necessary to have clear criteria for accepting the correct answers.

Input and the control of input. After the scoring, the information was input into the data base according to strictly set criteria, and then additionally checked in order to avoid, that is, to correct possible mistakes in the input.

4. BASIC PRINCIPLES OF SCORE INTERPRETATION FOR SUBJECT DOMAINS

Sanja Fulgosi

Results of testing of selected subject domains of Croatian language and Maths are given in detail in separate chapters of this *Report*.

Results of the *Project* given in this *Report* refer to the tests used to test particular subject domains, i.e. to the items testing particular subject domains.

The descriptions of *particular subject domains* with the help of the estimates of student achievements show the results of each item and its characteristics whose parameters can be used to describe the efficacy of both the testing and the whole *subject domain*.

The test for this testing has been designed according to *Item Response Theory* and it is therefore crucial to initially list the basic parameters on which score interpretation is generally based. These basic parameters are described with a sample graph that relates to score interpretation for both subjects. Two chapters point out to specific features related to approach to score interpretation for Croatian language and Maths.

Graph 1 shows the item range on the competence scale for each of the subject domains.

Estimates of student achievement are given on the x-axis, and individual items are given on the y-axis.

Estimates of student achievement are given in the range from 100 – the minimum achievement level, to 400 – the maximum achievement level with the central value of 250.

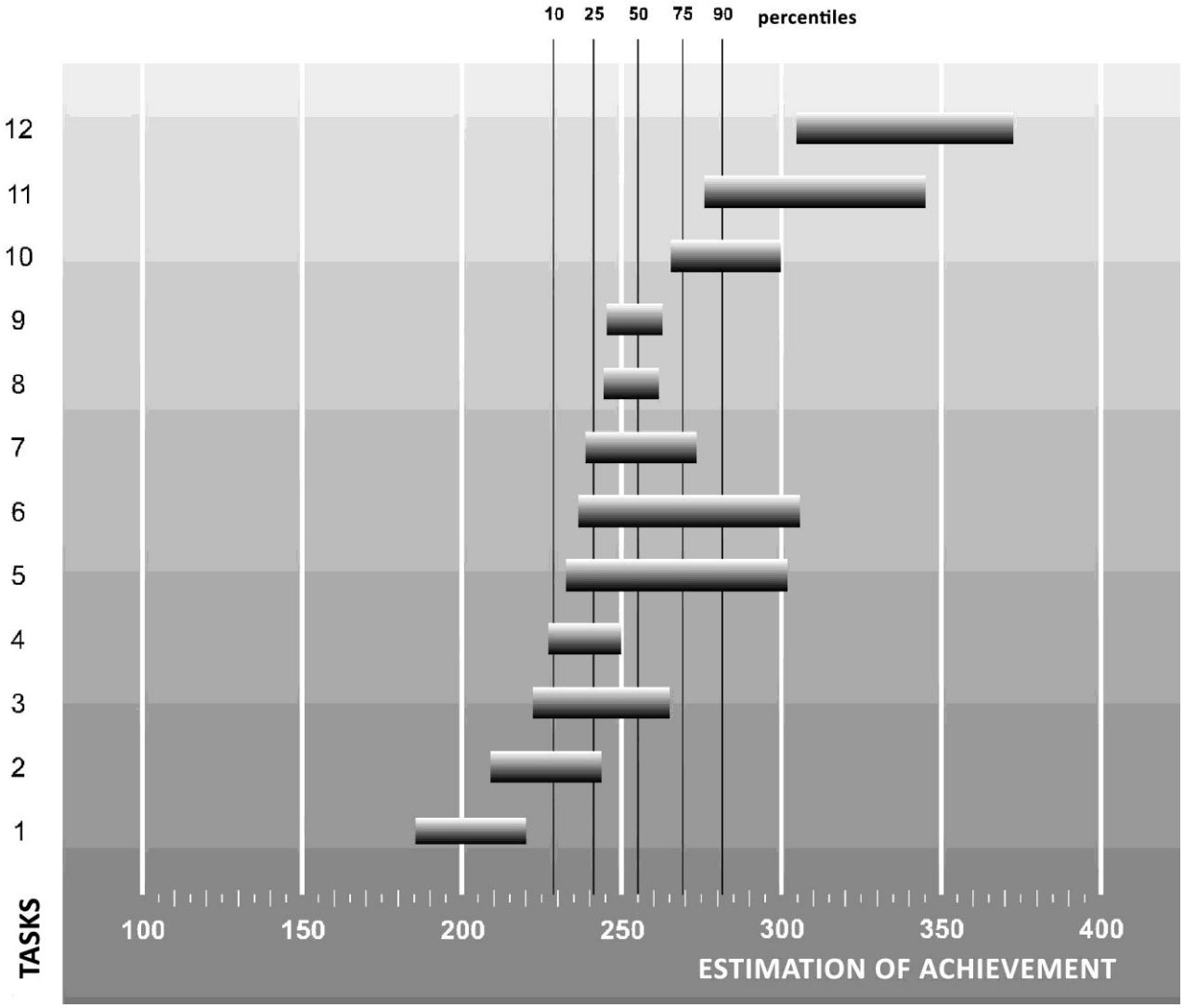
Estimating achievement refers to the estimated level of knowledge in particular *subject domain* based on the obtained scores, i.e. the accuracy of doing an item from the particular *subject domains*.

The items are given on the y-axis according to the gained difficulty estimate within the testing sample.

The difficulty level of each item is shown by its position within the area of estimated achievements that it covers. The bar length for each item shows its discrimination, i.e. the extent to which the item distinguishes between candidates within the competence scale. For the area of the estimated achievement that the bar length covers can be said that the item is moderately difficult. The starting point of the bar, on the left-hand side, is on the level of the estimated achievement at which the students with that level of the estimated achievement are 50% likely to get the item right. The end of the bar for each item is on the level of the estimated achievement at which the students with that level of estimated achievement have 80% chance of getting the item right. For all levels of estimated achievement left from the starting point of the bar item there is a high probability that they will do the item right, i.e. that the item will be easy for them, whereas for the levels of the estimated achievement right from the end point of the bar there is a very law probability that they will get the item right.

The value of test items is determined by their difficulty level and discrimination index. A test should contain equal representation of items of all levels distributed across the whole competence scale, and discriminative enough to distinguish within particular estimates of achievement.

Graph 1. Graph as an example of describing basic principles of score interpretation



Some of the obtained percentiles (10, 25, 50, 75, 90) are shown vertically on the scale.
Percentiles show the obtained sample distribution, i.e. they show what percentage of student sample in the displayed testing area attains particular levels of achievement. If we take percentile 10th as an example, we can see that if falls on the 228,7 achievement level, which means that 90% of student sample scored over 228,7, and 10% scored less than 228,7.

A normal distribution in the sample would assume the following distribution of percentiles: the 10th percentile rank is 186, the 25th percentile rank is 216, the 50th percentile rank is 250, the 75th percentile rank is 284, the 90th percentile rank is 314 on the competence scale.

The assumption of normal distribution of population's competence requires equal coverage of all areas of the competence scale with test items and an adequate representation of the number of

items and of item's characteristics in a particular range of competences.

4.1. CROATIAN LANGUAGE

The interpretation of scores from the subject of Croatian language is based on classifying values on the competences scale in three groups in each subject domain. These groups are determined in view of the range of items on the competences scale, i.e.in view of the ration of item length and the accompanying section of the sample. Hence three competence groups have been identified with the three groups of items:

a) easy items – answered correctly by 75% of the sample students

b) items of medium difficulty – answered correctly by 50% of the sample student

c) difficult items - answered correctly by 25% of the sample student

Score interpretation with the values on the scale is illustrated with particular examples of items and educational outcomes that refer to this competence according to this testing.

4.2. MATHS

Competence scale in the subject domains of Maths is described with respect to five groups of students, those on the 10th, 25th, 50th, 75th and 90th percentile, in three categories. It can generally be said that a student:

a) handles a particular item well (concept, skill) if there is more than 80% probability to do the item correctly (right-hand end of the bar length describing the item)

b) partially handles an item if this probability is between 50% and 80%

c) does not handle an item if there is less than 50% probability to do the item correctly (left-hand end of the bar length describing the item)

5. CROATIAN LANGUAGE - SUBJECT DOMAINS

Sanja Fulgosi

5.1. THEORY OF LITERATURE

One of the main guidelines in forming and setting objectives of testing Croatian language was the selection of the content whose basic terms are taught in lower grades of elementary school and which is later systematically further developed in the Curriculum for Gymnasium High School. This is emphasized in the chapter that deals with the selection of content.

One of such areas is the *theory of literature*. Considering the representation and the importance of this area for understanding literary texts, the aim was to determine to what extent, on a conceptual level, basic literary theoretical terminology outlined in the *Curriculum for Secondary Education*, for Croatian language have been learned in the first and second grade. The content of the testing referred to three literary theoretical areas:

a) theory of verse

b) figures of speech

c) characteristics of literary forms and genres.

Subject domain of the *theory of literature* was tested with fifteen objective type items: ten open-ended items (short-answer questions), and five closed (multiple-choice questions). Although, according to test specifications, *theory of literature* and *reading literary texts* form two different testing domains both can be tested using the same texts. Two texts, *The Sonnet 18 by* W. Shakespeare[7] and the fable *Tužno primirje by* Gustav Krklec[8], were used for testing both domains, *theory of literature* and *reading literary texts*: The text *Ribareva Jana* by August Šenoa[9] was used to test only the content of the *theory of literature*.

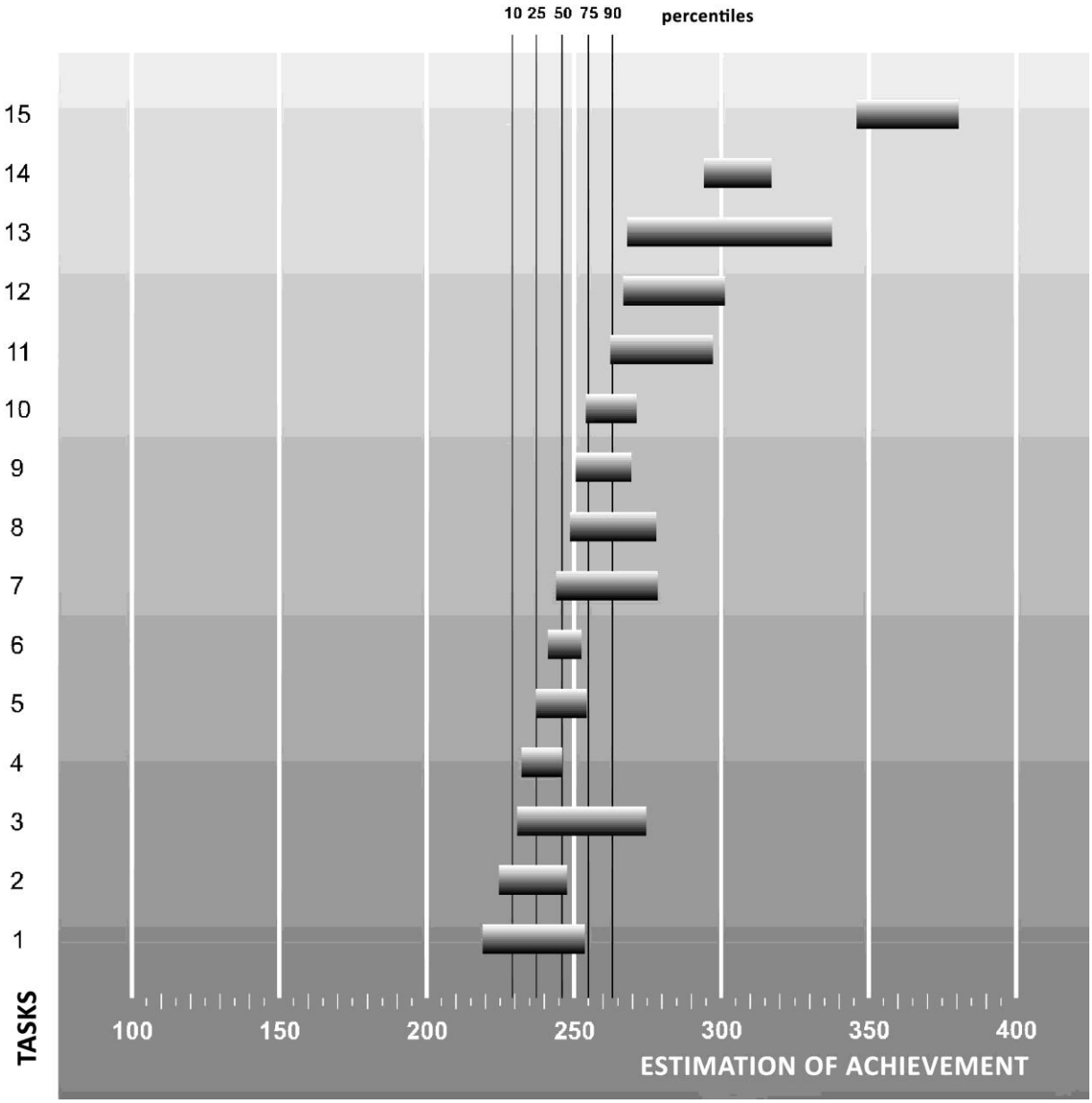
[7] adapted from William Shakespeare's *Sonnets* (translated by Luko Paljetak), Znanje, Zagreb, 1984
[8] adapted from Gustav Krklec, *Telegrafske basne*, Školska knjiga, Zagreb, 2001
[9] taken fromAugust Šenoa, *Budi svoj, pjesme i povjestice,* Nart- trgovina, Zagreb, 1999

5.1.1. Results in the subject domain of the theory of literature

The test results in the domain of the *theory of literature* are shown in *Graph 2* as the range of items from that domain on the competence scale.

Items on the scale can be divided into three groups according to their difficulty level: *easy, medium and difficult.*

Graph 2. Range of items from the domain of the *theory of literature*



The coverage curve of the domain of the *theory of literature* with items used in this testing shows at the same time low coverage of lower values of the competence scale in the sample (up to 25th percentile rank, i.e. up to the level of achievement evaluation in values from 100 to 237,25) and in the assumed normal distribution of population's competence (up to 25th percentile, i.e. up to the level of achievement evaluation in values from 100 to 216).

Representation of items in the middle part of the competence scale of the tested domain shows a good coverage on the scale which refers to the testing sample (from 25th to 75th percentile rank, i.e. up to the level of achievement evaluation in values from 237,25 to 255,05). This assumption also refers to the assumed normal distribution of a population's competence (from 25th to 75th percentile rank, i.e. up to the level of achievement evaluation in values from 216 to 284).

Representations of difficult items of the tested domain in this testing shows a relatively good coverage on the competence scale in view of the testing sample (above 75th percentile rank, i.e. up to the level of achievement evaluation in values from 255,05 to 400). However, in view of the assumed normal distribution of population's competence this representation is low (above 75th percentile rank, i.e. up to the level of achievement evaluation in values from 284to 400). Apart from the fact that the items cover the competence scale relatively poorly in higher values, there are also some parts of the competence scale that are not represented with items at all (between values 337,5 and 345,95).

There are major differences between the obtained distribution of the estimates of achievements and the assumed normal distribution of population's competence.

Items that are done correctly by 75% of sample students, i.e. items that in their range stretch from the values of 100 on the competence scale up to 25th percentile rank, are considered to be easy. Since there are no items in this research that fully meet this criterion, the items that are listed as easy will be used as an example for easy items. These items cover the lower part of the obtained scale. Taking *Graph 2* into consideration, items 1, 2, 3 and 4 can be considered to be easy. They can be used to estimate lower achievements, i.e. lower level of knowledge in the domain of the theory of literature, which is between the values 219,1 and 246,1 on the competence scale. Furthermore, *Graph 2* shows us that despite the fact that some items overlap in values on the competence scale, they are still part of the test because of the content value in the domain of the theory of literature.

The above mentioned competence range refers to the ability to identify basic elements of the literature in verse.

To students who reach estimated achievement on the 25th percentile rank in the obtained sample (237,25), items described as easy, following the criterion that 75% of testees determined by the 25th did the item correctly, are in fact of medium difficulty, which depends on the cutting score of each item. None of these items has a probability of over 80% to be done correctly by students who reach estimated achievement on the 25th percentile rank.

Items 1 and 3 have a very low discrimination index because their bars stretch across most of the obtained distribution of the sample of estimated achievement. These items have been kept only because of their content value.

The item given in *Example 1* belongs to the group of easy items.

Example 1. An easy item from the domain of *theory of literature*

TEXT - RIBAREVA JANA

What type of rhyme scheme is there in the second stanza of the attached text?

a) *couplet*

b) *monorhyme*

c) *split rhyme*

d) *four-line stanza*

e) *enclosed*

Item type: *multiple choice*

Level of difficulty: *easy*

Educational outcome: *conceptual knowledge; identifying and comprehending - identifying the rhyme type*

Key: *c)*

Competencies that 50% sample students have can be determined by items of medium difficulty, and they are between the 25th and the 75th percentile rank. According to this criterion based on the obtained outcomes, in view of distribution of the testing sample, only items 5 and 6 can be described as items of medium difficulty. Due to a very small distribution of the testing sample on the competence scale and to the descriptive value of other items, more items can be described as items of medium difficulty. Items of medium difficulty numbered 5, 6, 7, 8, 9 and 10 can be used to estimate achievements between the values 236,95 and 271,5 on the competence scale.

The above mentioned competence range refers to the ability to classify most of the literature forms in view of genres, understanding the characteristics of major forms, identifying and comprehending the structure of the basic types of stanza and verse in the given text.

To students who reach estimated achievement on the 50th percentile rank in the obtained sample, i.e. the value of 246,15, items 5, 6, and 7 are of medium difficulty depending on the cutting score of each item. For these items it can be said that there is a 50 - 80% probability that the students who reach the estimated achievement on the 50th percentile rank in the obtained sample will get the item right. For those students, items 8, 9 and 10 are difficult items since there is a low probability, less than 50%, that these students will get these items right.

As depicted by *Graph 2,* items 7 and 8 have a very low discrimination index because their bars stretch across the bigger part of the obtained distribution of the sample of estimated achievement.

Item 10 is described as an item of medium difficulty due to the lack of easy items although considering its range it should be a difficult item.

Items done correctly by 25% sample students, i.e. items that in their range stretch above the 75th percentile rank, are considered to be difficult. *Graph 2* shows us that those are items 11, 12, 13, 14 and 15. These items can be used to estimate the knowledge in the domain in the theory of literature in values above 255,05 on the competence scale.

Competence in values above 255,05 in this domain refers to identifying and understanding more complex figures of speech and determining genres of unfamiliar texts that by their structure and characteristics are not typical representatives of their category. A difficult item in this domain would be item 15 as shown in Graph 2.

To students who reach estimated achievement on the 75th percentile rank or lower in the obtained sample (255,05), items described as difficult items are indeed difficult. For these items it can be said that there is less than 50% probability that the students who reach the estimated achievement on the 75th percentile rank in the obtained sample will get the item right. In other words, 75% of students most probably will not get the item right.

Example 2. A difficult item from the domain of theory of literature

TEXT - RIBAREVA JANA

What is the literary form of the given text? _____

Item type: *short-answer question*

Level of difficulty: *difficult*

Educational outcome: *conceptual knowledge; understanding and applying - determining the literary form of a text whose structure and features are not typical for the given major form*

Key: *ballad*

Marking scheme:

CODE 1 - Answers directly stating that the major form in questions is **ballad.**

CODE 0 – Answers referring to other epic genre are not taken into consideration

Examples of incorrect answers: lyric, a lyric poem, epic genre, drama, epic, a love poem

5.2. READING LITERARY TEXTS

Considering the number of required literary works, the purpose of the lessons themselves as well as the way that literature pervades all other aspects of the Croatian language teaching, literary texts are the core of *Curriculum for Secondary Education* for Croatian language. At the same time, understanding different forms of non-literary forms enables the students to communicate more efficiently in all communication functions of Croatian language as already pointed out in the chapter that explains the testing content.

Due to all of the above mentioned facts, reading comprehension is tested with two subject domains: *reading literary texts* and *reading non-literary texts*. The testing frame and designing the items are the same in both subject domains that refer to reading. In this research two basic processes have been estimated, and reading comprehension items have been designed considering two levels.

1. finding certain information in a text:
    a) finding one or more pieces of information that can be found directly in the text
    b) finding one or more pieces of information that, at first sight, could meet more than one criterion from the text
    c) finding or setting relationships between information implied in the text

2. interpreting a text:
    a) identifying the topic of a text and the author's message
    b) identifying the main idea of a text and interpreting the relationship between different parts of the same text
    c) identifying and interpreting the idea of particular text parts which have a more complex relationship with the whole text that, at the first sight, could meet more than one criterion or that could be contradictory to the text as a whole.

The part of the text referring to the subject domain of *reading literary texts* consists of fifteen questions: ten closed (multiple-choice questions), and five open-ended items, three out which are short-answer questions and two are open-ended questions.

Reading comprehension items in the domain of *reading literary t*exts are designed based on four literary texts. Bearing in mind the literary works listed in the *Curriculum for Secondary Education* for Croatian language for the second grade, all chosen literary texts are to students unfamiliar literary texts. To test reading comprehension of literary texts the following texts have been chosen and adapted:

1. Williame Shakespeare – *Sonnet 18* [10]

2. Gustav Krklec – *Tužno primirje*[11]*,* fable in verse

3. Ranko Marinković (2002) – *Don Kuzmino uho*[12]*,* a paragraph taken from the novel *Kiklop*

4. *Gilgameš*[13], a paragraph from the epic.

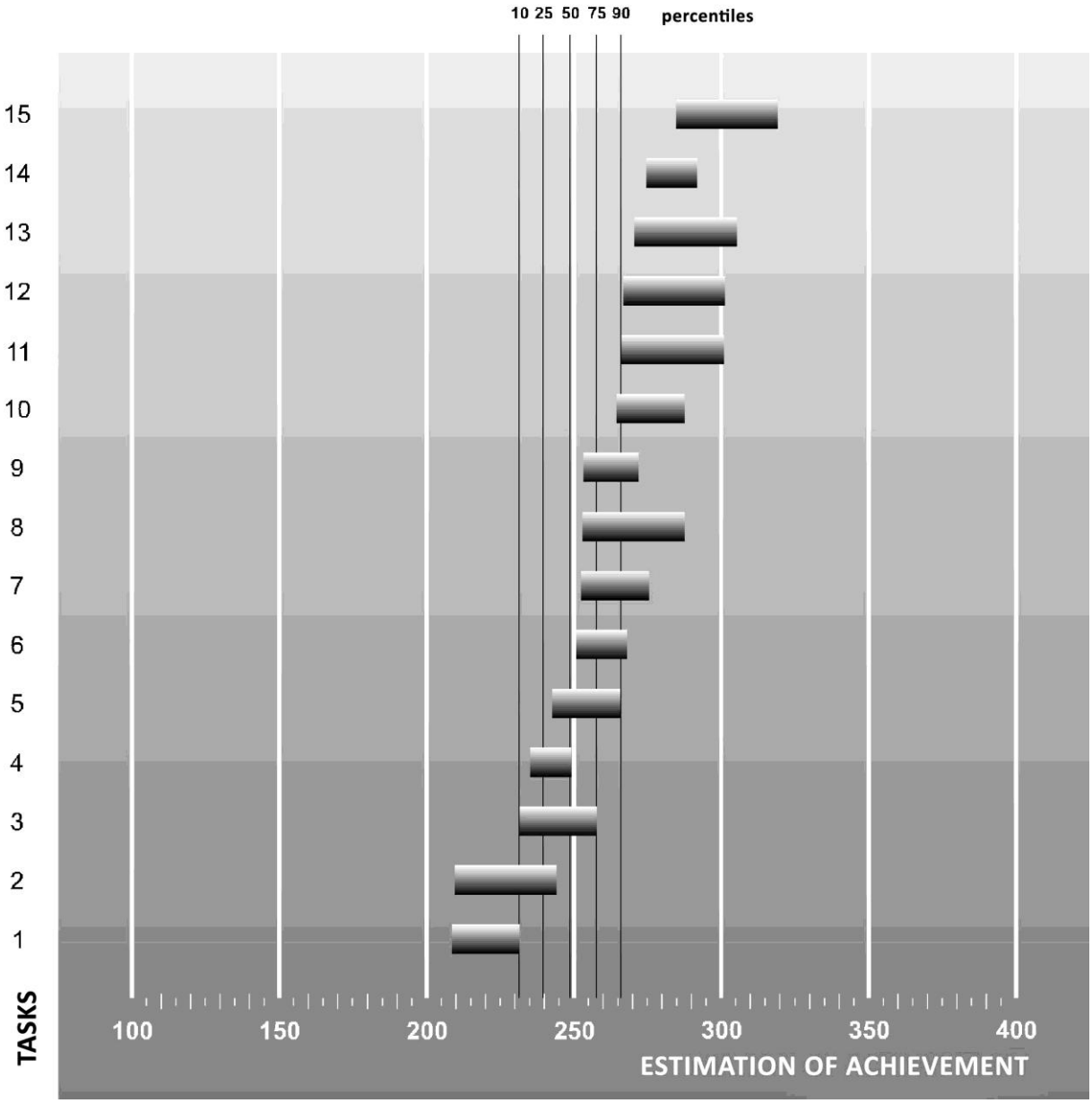
[10] adapted from William Shakespeare's *Sonnets* (translated by Luko Paljetak), Znanje, Zagreb, 1984
[11] adapted from Gustav Krklec, *Telegrafske basne*, Školska knjiga, Zagreb, 2001
[12] taken from Ranko Marinković, *Kiklop*, Otokar Keršovani, Rijeka, 2002
[13] taken from Nevenka Košutić – Brozović*, Čitanka stranih književnosti,* Školska knjiga*,* Zagreb, 1970

5.2.1. Results in the subject domain of reading literary texts

Test results in the domain of *reading literary texts* are shown in *Graph 3* as the range of items from that domain on the competence scale.

Items on the scale can be divided into three groups according to their difficulty level: *easy, medium and difficult.*

Graph 3. Range of items from the domain of *reading literary texts*



The coverage curve for the domain of *reading literary texts* with items used in this testing shows at the same time low coverage of lower values of the competence scale in the sample (up to the 25th, i.e. up to the level of achievement evaluation in values from 100 to 239,7) and in the assumed normal distribution of population's competence (up to 25th percentile, i.e. up to the level of achievement evaluation in values from 100 to 216).

Representation of items of medium difficulty of the tested domain in this testing shows a good coverage on the competence scale both in the testing sample (from the 25th to the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 239.7 to 257.8) and in the assumed normal distribution of a population's competence (from the 25th to the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 216 to 284).

Representations of difficult items of the tested domain in this testing shows a relatively poor coverage on the competence scale in view of the testing sample (above the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 257,08 to 400, and in view of the assumed normal distribution of population's competence this coverage is extremely low (above the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 284 to 400).

There are major differences between the obtained distribution of the estimates of achievements and the assumed normal distribution of population's competence.

Items that are done correctly by 75% of sample students, i.e. items that in their range stretch from the values of 100 on the competence scale up to the 25th percentile rank, are considered to be easy. Since in this research only item 1 fully meets this criterion, other items that are listed as easy will be only used as an example for easy items. These items cover the lower part of the obtained scale. Taking *Graph 3* into consideration we can say that items 1, 2, 3 and 4 can be considered to be easy. They can be used to estimate lower achievements, i.e. lower level of knowledge in the domain of reading literary texts which is between the values 208,7 and 249,1 on the competence scale. Furthermore, *Graph 3* shows us that despite the fact that some items overlap in values on the competence scale, they are still part of the test because of the content value in the domain of reading literary works.

The above mentioned competence range refers to simple reading comprehension items which ask a student to locate information explicitly stated in the text and interpret the main idea of the text. An example of an easy reading comprehension item is given *in Example 3*.

To students who reach estimated achievement on the 25th percentile rank in the sample (239,7), items 2, 3 and 4 are of medium difficulty, which depends on the cutting score of each item. Item 1 is easy and it is the only one that has a probability of over 80% to be done correctly by students who reach estimated achievement on the 25th percentile rank in the testing sample.

Example 3. An easy item from the domain of *reading literary texts*

TEXT - GILGAMEŠ

What is the topic of the given text?

a) *conflict between gods and people*

b) *heroic actions*

c) *death and eternity*

d) *conflict between two friends*

Item type: *multiple choice*

Level of difficulty: *easy*

Reading comprehension item: *interpreting a text - identify the topic and the author's intention*

Key: *a)*

Competencies that 50% sample students have and whose estimated achievement is between the 25th to 75th percentile rank, can be determined by five items of medium difficulty, and these items range from the 25th to the 75th percentile rank. None of the items fully meets this criterion. Furthermore, there is a very small distribution of the testing sample on the competence scale and the discrimination index of some items is very low. Due to this as well as to the earlier adaption of criterion in the need of more easy items, items described as items of medium difficulty can only be considered to be an example. Items of medium difficulty numbered 5, 6, 7, and 9 can be used to estimate achievements between the values 242,75 and 272 on the competence scale.

To students who reach estimated achievement on the 50th percentile rank in the obtained sample, i.e. the value of 248,75, all of the items described as items of medium difficulty (apart from item 5) are difficult items. For these items it can be said that there is less than 50% probability that the students who reach the estimated achievement on the 50th percentile rank in the obtained sample will get the item right.

Item 5 is for those students an item of medium difficulty. For that item it can be said that there is 50-80% probability that the students who reach the estimated achievement on the 50th percentile rank in the obtained sample will get the item right.

Items done correctly by 25% sample students, i.e. items that in their range stretch above the 75th percentile rank, are considered to be difficult. Graph 3 shows us that those are items 10, 11, 12, 13, 14 and 15. These items can be used to estimate the knowledge in the domain of *reading literary texts* in values above 264,65 on the competence scale.

Difficult items in the domain of reading literary texts roughly refer to finding information implied in the text, recognising the idea of particular text parts that have a complicated relationship with the text as a whole. An example of a difficult item in the domain *reading literary texts* as well as the rating guidelines is given in

Example 4.

To students who reach estimated achievement on the 75th percentile rank or lower in the obtained sample (257,8), items described as difficult items are indeed difficult. For these items it can be said that there is less than 50% probability that the students who reach the estimated achievement on the 75th percentile rank in the obtained sample will get the item right. In other words, 75% of students most probably will not get the item right.

Example 4. A difficult item from the domain of *reading literary texts*

TEXT - TUŽNO PRIMIRJE

What is the main idea of the given text? Answer in one sentence.

_____

Item type: *open-ended question*

Level of difficulty: *difficult*

Educational outcome: *interpreting a text - identifying and understanding the main idea of a text part which has a complex relationship with the whole text*

Marking scheme:

Answers containing content elements as follows are considered to be correct:

CODE 1:

- A wolf and a lamb can never be friends.
- A wolf will never change.
- A wolf will always eat a lamb.
- directly stated opposite (strong - weak or naive - cunning) that will never be overcome
- emphasised eternal conflict between the naive and the cunning
- the relationship do not change, the tragic end is predictable
- there is no possibility of closeness
- inevitable tragic end for the naive and weak
- answers that mention the eternal opposition and the tragic end for the weak and naive (often, perhaps, it could happen, etc.)

Examples of correct answers (in accordance with CODE 1):

The main idea is: *two enemies can never make a truce because one of them will always have evil intentions.*

The cunning ones will always take advantage of the naive for their own benefit.

Incorrect answers are all that do not match the contents described in CODE 1 or they partially match. Examples of incorrect answers:

The main point is that a truce should be made.

Conflicts of any kind should be avoided.

One can always find a real friend.

The ability to find information, interpret the meaning and the relationship between pieces of information in texts that students come across on daily basis is crucial for a communication in Croatian language as well as for everyday life. As already pointed out, in both content areas referring to reading, reading comprehension was be tested considering in regard to two basic processes, and reading comprehension items have been designed considering the two levels.

1. finding certain information in a text:

    a) finding one or more pieces of information that can be found directly in the text

    b) finding one or more pieces of information that, at the first sight, could meet more than one criterion from the text

    c) finding or setting relationships between information implied in the text

2. interpreting a text:

    a) identifying the topic of a text and the author's message

    b) identifying the main idea of a text and interpreting the relationship between different parts of the same text

    c) identifying and interpreting the idea of particular text parts which have a more complex relationship with the whole text that, at first sight, could meet more than one criterion or that could be contradictory to the text as a whole.

Texts used to test the domain of reading a non-literary text have been chosen in accordance with two criterion:

a) the text's structure - uninterrupted text and interrupted text (a table and a form)
b) the text's topic - educational, public and scientific.

The part of the text that refers to the subject domain of *reading non-literary texts* consisted of 19 objective type questions: ten closed (multiple-choice questions), and nine open-ended items, two out which are open-ended questions and seven are short-answer questions.

To test reading comprehension of non-literary works the following texts have been chosen and adapted for this testing:

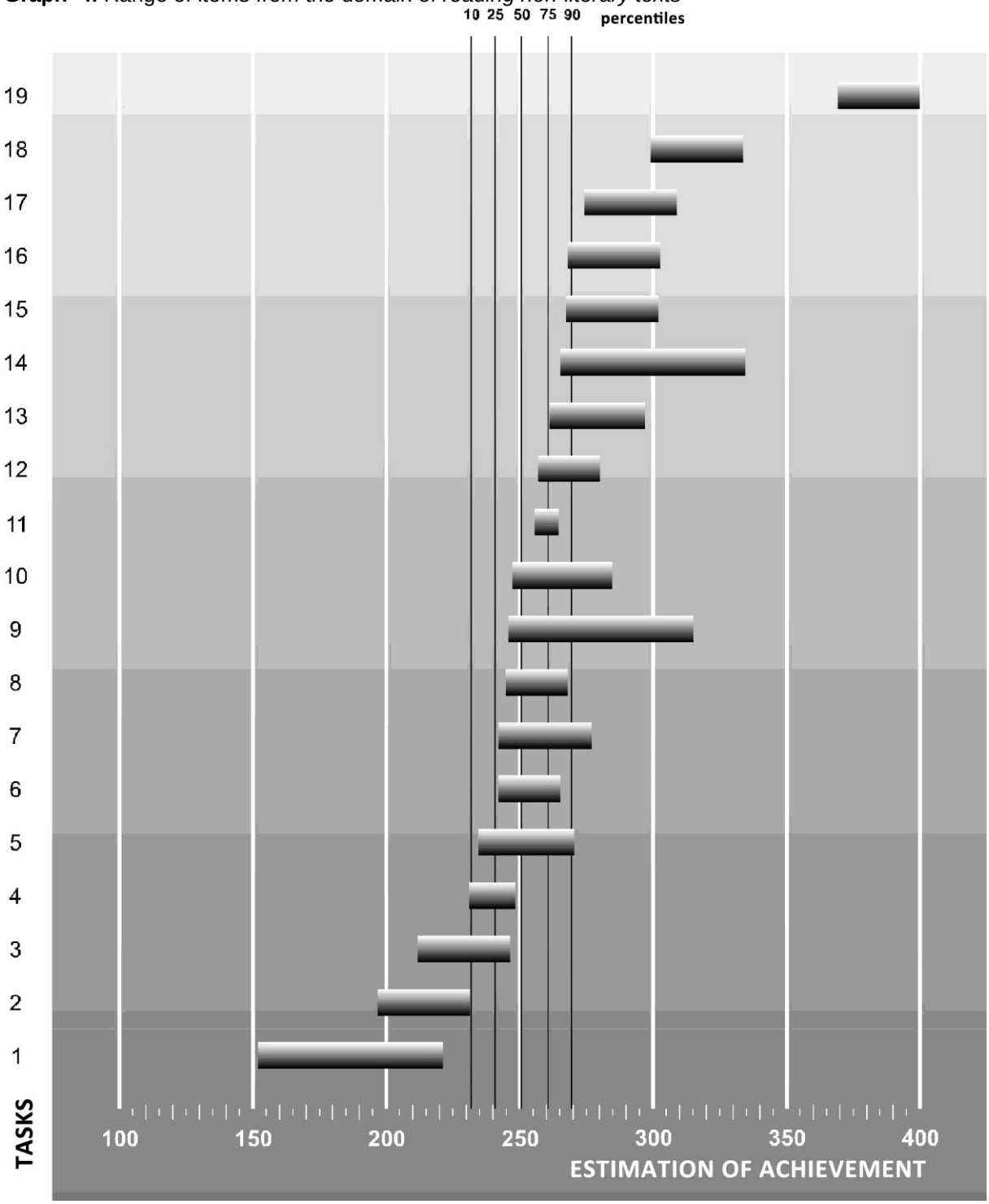
1. *Horizontal and vertical cultures*[14]
2. *Package tours contract for end-of-school trip*[15]
3. *The Olympic Games*[16]
4. Admission form[17].

[14] taken and adapted from Nives Opačić, *Hrvatski u zagradama*, Hrvatska sveučilišna naklada, Zagreb, 2006, Aleksandar Flaker *Proza u trapericama*, SNL, Zagreb, 1983
[15] PACKAGE TOURS CONTRACT FOR END-OF-SCHOOL TRIP – adapted from the contract downloaded from the following website:
http://www.speranza.hr/documents/UVJETI_PUTOVANJA_MATURALAC_2008.pdf on 20th July 2008
[16] OLYMPIC GAMES - adapted from the texts downloaded from the following websites:
http://hr.wikipedia.org/wiki/Olimpijske_igre/Starovjekovne_Olimpijske_igre
http://hr.wikipedia.org/wiki/Olimpijci_s_najviše_medalja 20th July 2008
[17] adapted from the form downloaded from the following website:
http://www.dphj.hr/dokumenti/pristupnica%20u%20DPHJ%20-%20obrazac.pdf 20th July 2008

5.3.1. Results in the subject domain of *reading non-literary tex*ts

Test results in the domain of *reading non-literary texts* are shown in *Graph 4* as the range of items from that domain on the competence scale.

Graph 4. Range of items from the domain of *reading non-literary texts*



The difficulty of the items on the competence scale can be seen in *Graph 4*, i.e. their estimated achievement level in the subject domain of *reading non-literary texts*. Items on the scale can be divided into three groups according to their difficulty level: *easy, medium and difficult.*

The coverage curve for the domain of *reading non-literary texts* with items used in this testing shows a relatively good coverage of lower values of the competence scale in the obtained testing sample (up to the 25th percentile rank, i.e. up to the level of achievement evaluation in values from 100 to 240,8).This representation is poor considering the assumed normal distribution of population's competence (up to 25th percentile, i.e. up to the level of achievement evaluation in values from 100 to 216).

Representation of items of medium difficulty of the tested domain in this testing shows a good coverage on the competence scale both in the testing sample (from the 25th to 75th percentile rank, i.e. up to the level of achievement evaluation in values from 240,8 to 260,6) and in the assumed normal distribution of a population's competence (from the 25th to 75th percentile rank, i.e. up to the level of achievement evaluation in values from 216 to 284).

Representations of difficult items of the tested domain in this testing shows a good coverage on the competence scale in view of the testing sample (above the 75th percentile rank, i.e. above the level of achievement evaluation in values from 260,6 to 400), and in view of the assumed normal distribution of population's competence this coverage is relatively good (above the 75th percentile rank, i.e. above the level of achievement evaluation in values from 284 to 400). There is no continuity of the coverage of the scale of estimated achievements of values between 333,8 and 369,25.

There are major differences between the obtained distribution of the estimates of achievements and the assumed normal distribution of population's competence.

Items that are done correctly by 75% of sample students, i.e. items that in their range stretch on the competence scale up to the 25th percentile rank, are considered to be easy. Items 1 and 2 fully meet this criterion. Since there are not many items with lower values on the scale of estimated achievement, the above mentioned criterion will be implied as to have a greater number of easy items. Taking *Graph 4* into consideration we can say that items 1, 2, 3 and 4 can be considered to be easy. They can be used to estimate lower knowledge in the domain of *reading non-literary texts* which is between the values 151,85 and 248,4 on the competence scale. Furthermore, Graph 4 shows us that despite the fact that some items overlap in values on the competence scale, they are still part of the test because of the content value in the domain of *reading non-literary texts*.

To students who reach estimated achievement on the 25th percentile rank in the obtained sample (240,8), items 1 and 2 are easy because there is a probability of over 80% that they will do them correctly. Items 3 and 4 are of medium difficulty since there is a 50-80% probability that they will be done correctly by students who reach estimated achievement on the 25th percentile rank depending on the cutting score of each item.

The above competence range refers to simple reading comprehension items which ask a student to locate information explicitly stated in the text and interpret the main idea of the text. An example of an easy reading comprehension item is given in example 6.

Example 6. An easy item from the domain of *reading non-literary texts*

TEXT - OLYMPIC GAMES

According to Table 1, write the number of athlete who has the longest history of taking part in the Olympics.

_____

Item type: *short-answer question*

Level of difficulty: *easy*

Reading comprehension item: *finding information - finding one or more pieces of information explicitly stated in the text*

Key: *6*

Marking scheme:

CODE 1: Answers containing 6 / 6th / six / sixth

Items of medium difficulty cover the scale of estimated achievement that refers to 50% sample students, i.e. items that range from the 25th to 75th percentile rank. Due to the very small distribution of the items of medium difficulty on the competence scale, none of the items in the testing sample fully meets this criterion. Due to this as well as to the earlier adaption of criterion, a wider range of items has been included. Therefore, items of medium difficulty numbered 5, 6, 7, 8, 9, 10, 11 and 12 can be used to estimate achievements between the values 234,6 and 280,1 on the competence scale.

The above mentioned competence range in the domain of *reading non-literary texts* refers to finding pieces of information that have a more complex relationship within the text as well as interpreting the idea of parts of the text. An example of a reading comprehension item of medium difficulty is given in *Example 7*.

To students who reach estimated achievement on the 50th percentile rank in the obtained sample (250,7), items 5, 6, 7, 8, 9, and 10 are items of medium difficulty depending on the cutting score of each item. For these items it can be said that there is 50-80% probability that the students who reach the estimated achievement on the 75th percentile rank in the obtained sample will get the item right. Items 11 and 12 are difficult since there is less than 50% probability that the students who reach the estimated achievement on the 50th percentile rank in the obtained sample will get the item right.

Items 5, 9 and 10 need further consideration due to low discrimination index, i.e. the bar length of the item range, with a relatively small range in the obtained results of the testing sample, does not classify them precisely in any of the item difficulty categories. The bar length of item 5 stretches from below the 25th percentile rank up to the 90th percentile rank (values of 234,6 to 270,55), i.e. it covers too wide range of estimated achievements in the testing sample as to be discriminative. Bar lengths of items 9 and 10 also cover too wide range of estimated achievements within the testing sample, i.e. discrimination index is too low. Despite of this, these items are still part of the test because of the content value in the domain of

reading non-literary texts.

Example 7. A difficult item from the domain of *reading non-literary texts*

TEXT - ADMISSION FORM

What is the aim of Vladimir Nazor's verse at the beginning of the Admission Form?

a) *encourage the fundamental purpose and tasks of the society*

b) *encourage the society members to read poetry*

c) *highlight Nazor as an important Croatian writer*

d) *encourage studying of Croatian language*

Item type: *multiple choice*

Level of difficulty: *medium*

Reading comprehension item: *finding information - identifying and interpreting the idea of a particular text part which has a more complex relationship with the whole text that, at first sight, could meet more than one criterion*

Key: *a)*

Items done correctly by 25% sample students, i.e. items that in their range stretch above the 75th percentile rank, are considered to be difficult. *Graph 4* shows us that those are items 13, 14, 15, 16, 17, 18 and 19. These items can be used to estimate the knowledge in the domain of *reading non-literary texts* in values above 261,05 on the competence scale.

To students who reach estimated achievement on the 75th percentile rank (260,6) or lower in the obtained sample, items described as difficult items are indeed difficult. For these items it can be said that there is less than 50% probability that the students who reach the estimated achievement on the 75th percentile rank or lower in the obtained sample will get the item right. In other words, 75% of students most probably will not get the item right.

5.4. GRAMMAR

Subject domain of *grammar* in this testing covered domains prescribed in the chapter Language of the *Curriculum for Secondary Education* for Croatian Language for the second grade. Within this subject domain it was tested to what extent, on the conceptual level, the knowledge of language use was acquired. This subject domain was not tested on one text, but most of the items were based on a language form found in a text adapted for the particular item. Subject domain of grammar refers to three areas:

a) basic linguistic forms

b) lexical category

c) grammatical category of word classes: nouns, pronouns, adjectives and numbers.

Subject domain of grammar was tested with twelve items: eleven closed (multiple-choice questions) and one open-ended item (short-answer questions).

5.4.1. Results in the subject domain of grammar

Test results in the domain of *grammar* are shown in *Graph 5* as the range of items from that domain on the competence scale.

Items on the scale can be divided into three groups according to their difficulty level: *easy, medium and difficult.*

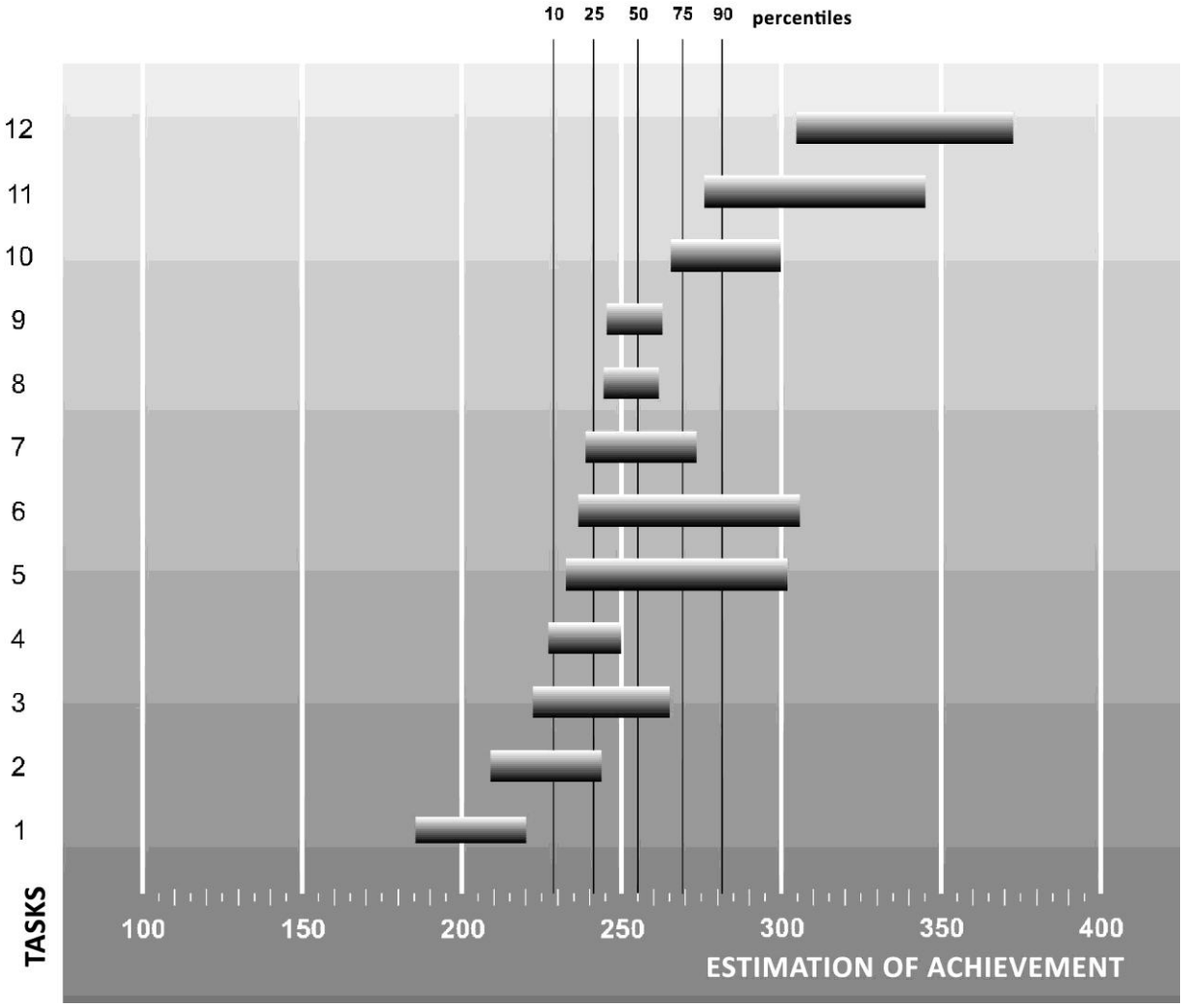
The coverage curve for the domain of *grammar* with items used in this testing shows low coverage of lower values of the competence scale in the testing sample (up to the 25th percentile rank, i.e. up to the level of achievement evaluation in values from 100 to 241,25). Coverage of the tested domain in the assumed normal distribution of population's competence (up to 25th percentile, i.e. up to the level of achievement evaluation in values from 100 to 216) can be described as extremely low.

Representation of items of medium difficulty in the tested domain in this testing there is a good coverage on the scale in the testing sample (from the 25th to the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 241,25 to 269,05). In the assumed normal distribution of a population's competence (from the 25th to the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 216 to 284)this is a good coverage.

Representation of difficult items of the tested domain in this testing shows a relatively good coverage on the competence scale in view of the testing sample (above the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 269,05 to 400). However, in view of the assumed normal distribution of population's competence this coverage is low (above the 75th percentile rank, i.e. up to the level of achievement evaluation in values from 284 to 400).

There are major differences between the obtained distribution of the estimates of achievements and the assumed normal distribution of population's competence.

Graph 5. Range of items from the domain of *grammar*



Items that are done correctly by 75% of sample students, i.e. items that in their range stretch below the 25th percentile rank, are considered to be easy. Only item 1 fully meets this criterion. Considering this, the low coverage of the lower part of the obtained scale and a relatively small number of items in this subject domain, this criterion had to be altered. Taking *Graph 5* into consideration items 1, 2, 3 and 4 can be considered to be easy. They can be used to estimate lower achievements, i.e. lower level of knowledge in the domain of grammar which is between the values 183,35 and 250 on the competence scale. Furthermore, Graph 2 shows us that despite the fact that some items overlap in values on the competence scale, they are still part of the test because of the content value in the domain of grammar.

The above competence range includes identifying and understanding of basic linguistic terms and identifying grammatical category of noun-word classes in a sentence. The item in example 7 belongs to the group of easy items.

To students who reach estimated achievement on the 25th percentile rank in the obtained sample (241,25), item 1 is easy as there is over 80% probability that they will do the item well. Items 2, 3 and 4 are for those students of medium difficulty since, depending on the cutting score of each item, there is a probability of 50- 80% these items will be done correctly by students who reach estimated achievement on

the 25th percentile rank.

Item 3 needs a further consideration due to its low discrimination index, i.e. the bar length of the item range, with a relatively small range in the obtained results of the testing sample, does not classify it to be an easy item. The bar length of item 3 stretches from below the 25th percentile rank up to over the 50th percentile rank (values of 222,2 to 265,05), i.e. it covers too wide range of estimated achievements in the testing sample as to be discriminative. Due to its of the content value item 3 is described as an easy item.

Example 7. An easy item from the domain of grammar

What is the term used for a word alternation regarding scaling of properties?

a) *declension*

b) *gradability*

c) *conjuction*

Item type: *multiple choice*

Level of difficulty: *easy*

Educational outcome: *factual knowledge - understanding - recognising basic linguistic forms*

Key: *b)*

Competencies that 50% sample students have and whose estimated achievement is between the 25th to the 75th percentile rank, can be determined by items of medium difficulty. Due to the very small distribution of the testing sample on the competence scale only items 8 and 9 fully meet the mentioned criterion. Considering the described value of the rest of the items, the small distribution as well as the criterion shift in the category of easy items, this criterion had to be broadened. Items numbered 5, 6, 7, 8, and 9 are considered to be items of medium difficulty and they can be used to estimate achievements between the values 232,6 and 305,85 on the competence scale.

The above competence range includes identifying and understanding of less typical word classes as well as the rules of declension carried out in accordance with more complex linguistic rules.

To students who reach estimated achievement on the 50th percentile rank in the obtained sample (255,15), all of the items described as items of medium difficulty are difficult items depending on the cutting score for each item. For these items it can be said that there is 50-80% probability that the students who reach the estimated achievement on the 50th percentile rank in the obtained sample will get the item right.

Item 7 has a low discrimination index since its bar length stretches from below the 25th percentile rank up to over the 75th percentile rank thus covering a bit more than the whole range of the items of the medium difficulty in the testing sample. In case it was removed from the category of items of medium difficulty, there would be a lack of coverage on the scale of estimated achievements between the value of 262,7 and 265,45.

Despite the fact that items 8 and 9 overlap in values on the competence scale, they are still part of the test because of the content value in the domain of grammar.

Items 5 and 6 need a further consideration due to their low discrimination index, i.e. the bar length of the

item range, with a relatively small range in the obtained results of the testing sample, does not allow their classification as items of medium difficulty. The bar length of items 5 and 6 stretches from below the 25th percentile rank up to over the 90th percentile rank, i.e. they cover a very wide range of estimated achievements in the testing sample and their discrimination index is very low. These items have been used as examples and due to their content value.

Items done correctly by 25% sample students, i.e. items that in their range stretch above the 75th percentile rank, are considered to be difficult. *Graph 5* shows us that those are items 10, 11 and 12. Although item 10 does not fully meet this criterion of difficult items since its left-hand side end of the bar is below the 75th percentile rank, it will still be classified as a difficult item because the greatest part on the scale of estimated achievement is over the 75th percentile rank. These difficult items can be used to estimate the knowledge in the domain of *grammar* in values above 265,45 on the competence scale.

Difficult items in the domain of grammar refer to recognising and determining the word forms of a less typical category as well as understanding logical. An example of a difficult item in the domain reading a literary text as well as the rating guidelines is given in *Example 8*.

A difficult item refers to testing following skills: understanding and defining less typical word classes, as well as understanding rules for declension of word classes. The item in *Example 8* belongs to the group of difficult items.

To students who reach estimated achievement on the 75th percentile rank (269,05) or lower in the obtained sample, items 11 and 12 are difficult and for these items it can be said that there is less than 50% probability that these students will get the item right. In other words, 75% of students most probably will not get the item right. For item 10 there is about 54% probability that the students who reach the estimated achievement on the 75th percentile rank in the obtained sample will get this item right.

Example 8. A difficult item from the domain of *grammar*

A noun in Nominative case has the form ptica (bird), in Genitive ptice, and in Dative ptici. Pronoun ja (I) in Nominative case has the form ja, in Genitive mene, and in Dative meni. Which declension scheme is represented in both noun and pronoun?

a) a- declension

b) i- declension

c) e- declension

Item type: *multiple choice*

Level of difficulty: *difficult*

Educational outcome: *conceptual knowledge - understanding rules of declension in Croatian language*

Key: c*)*

6. MATHS – SUBJECT DOMAINS

Željka Milin Šipuš, Sanja Antoliš, Jelena Gusić, Dragica Martinović, Kristina Penzar, Josipa Pavlić, Ljubica Stanković, Jagoda Krajina, Dobrila Golubović, Miroslav Smuđ

6.1. FRACTIONS

Sanja Antoliš, Jelena Gusić, Željka Milin Šipuš

The subject domain of *fractions* was tested with 24 items. Due to a relatively small number of items required by the test structure, and because of the use of pocket calculator, this domain mostly tested calculating using algebraic fractions, with only one item with fractions (with numbers). The items were open type, except for one complex multiple choice item.

The items assessed the student's capability of:

- comparison of fractions
- multiplication and division of algebraic fractions and the presentation of the results in reduced form
- addition and subtraction of algebraic fractions
- calculating with algebraic fractions, respecting the order of operations
- cancellation of algebraic fraction
- calculating with dual algebraic fractions.

In the subject domain related to *fractions* there were no correlations to other subject domains. That means that in the items in this domain it was not necessary to implement algebraic formulas like the ones for the binominal power and power difference. For that reason, an item like $\frac{a(a-4)}{(a-4)^2}$ was selected and not $\frac{a^2-5a}{a^2-10a+25}$. The fact that the student can not successfully solve the item $\frac{a^2-5a}{a^2-10a+25}$ does not reveal the reason why. It is possible that the student did not recognise the appropriate method for solving the item (numerator and denominator factorization), or he was unsuccessful during the factorization, or did not know how to cancel an algebraic fraction. In case of a lack of success in solving a fraction like $\frac{a(a-4)}{(a-4)^2}$, we know that the student did not know how to cancel the fraction containing simple algebraic expressions. Power calculation was not assessed, except for the simple calculation with positive power, such as for example x^6. Also, calculations with brackets were not assessed.

Table 12. Items schedule according to groups in the *fractions* subject domain

Item description	Item sequence number on the graph
Comparison of fractions	1, 2, 21
Multiplication and division of algebraic fractions and presenting results in reduced form	3, 4, 5, 6, 7, 8, 13, 15, 22
Addition and subtraction of algebraic fractions	10, 12, 19, 20, 23
Calculating with algebraic fractions respecting the order of operations	14
Cancellation of algebraic fraction	9, 11, 15, 17, 18, 24
Calculating with dual algebraic fraction	16

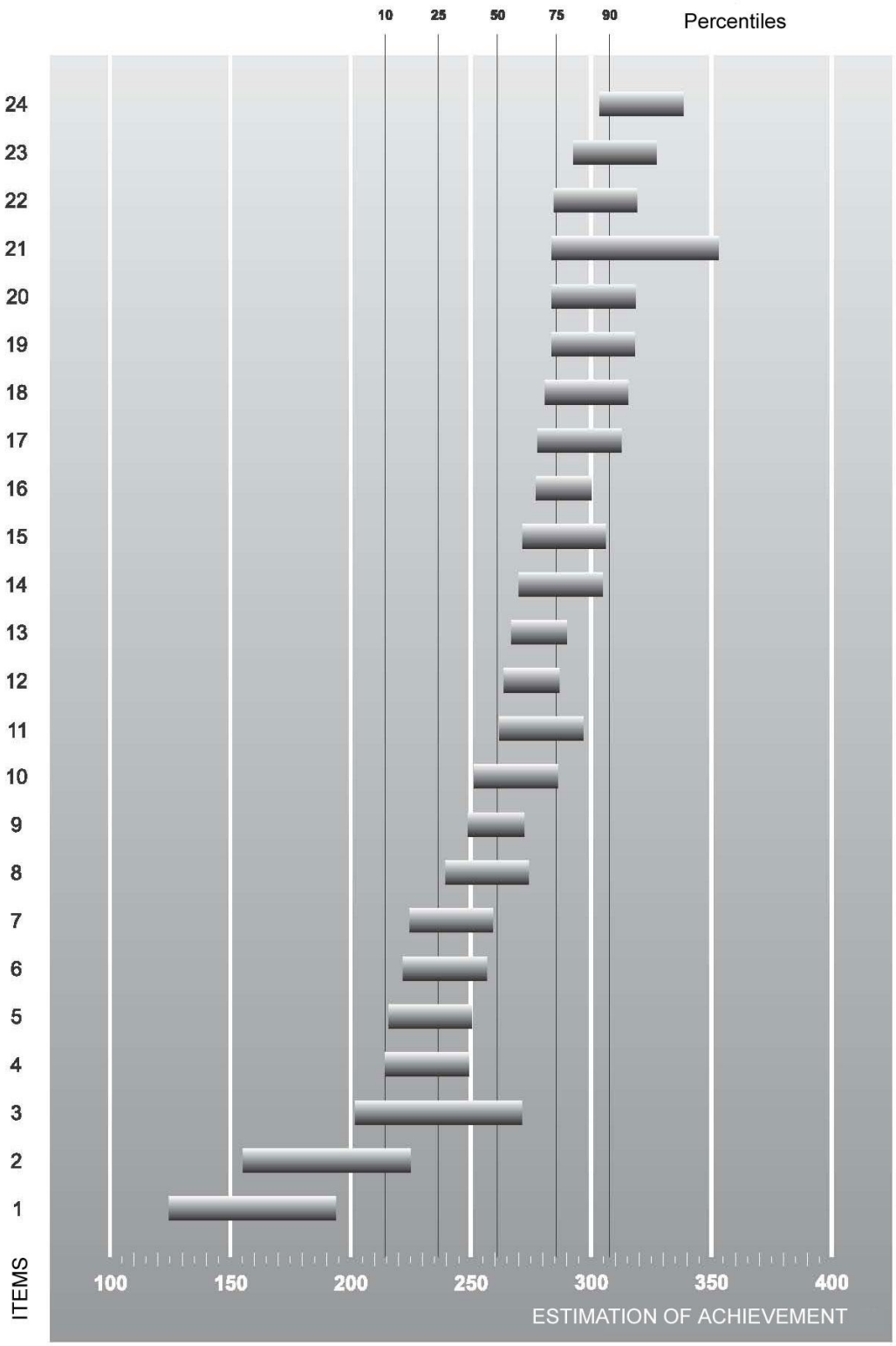
During the testing the correct item solving efficiency was between 24,1% and 91,1%. Some types of items, their sequence number regarding their difficulty, and solving efficiency percentage are listed in the following table.

Table 13. Item solving efficiency in the *fractions* subject domain

Item sequence number on the graph	Items	Item solving efficiency
4.	$\frac{x-2y}{4}\cdot\frac{8}{x-2y}$	78.9%
6.	$\frac{6a}{15}:\frac{3}{5}$	73.7%
9.	$\frac{a(a-4)}{(a-4)^2}$	62.8%

Item sequence number on the graph	Items	Item solving efficiency
12.	$\frac{5}{2a}+\frac{3+a}{a}$	44.8%
15.	$\frac{x-2y}{8}\cdot\frac{4}{2y-x}$	44.1%
17.	$\frac{a(a-4)}{(4-a)^2}$	36.3%
18.	$\frac{10x^7}{2x^3(x+1)}$	35.0%
19.	$\frac{5}{2a}-\frac{3+a}{a}$	35.9%
24.	$\frac{10x^7}{2x^4+2x^3}$	24.1%

Graph 6. Item range in the *fraction*s subject domain



Student on the 10th percentile rank manages to solve only the first item. This is the item where it is required to do a simple fraction extension. He can partially solve the second and third items. In those items, comparing fractions and dividing numbers by fractions are tested. That student does not manage to solve any of the Table12 items. These items are too difficult for him. However, the probability of his solving the item four is only $\frac{x-2y}{4}\cdot\frac{8}{x-2y}$ tly lower than 50%.

Student on the 25th percentile rank manages to solve the first and second item from the *fraction* subject domain. These are the items in which it is necessary to do simple comparisons and extensions of fractions. He partly manages to solve the items from three to seven, which test multiplication, division and cancellation like in the test items $\frac{6a}{15}:\frac{3}{5}$ or $\frac{x-2y}{4}\cdot\frac{8}{x-2y}$. He does not manage to solve all the other items.

Student on the 50th percentile rank manages to solve the first seven items (except item three which he manages to solve partially), which test fraction comparison, simple fraction extension, multiplication, division and cancellation like in item $\frac{6a}{15}:\frac{3}{5}$ or in item $\frac{x-2y}{4}\cdot\frac{8}{x-2y}$ .He partially manages to solve items eight to ten. These items test multiplication, division and cancellation of fractions which comprise simple terms like in item $\frac{a(a-4)}{(a-4)^2}$ and addition of fractions and numbers. He does not manage to solve all the other items. For example, the probability to fully solve item number twelve $\frac{5}{2a}+\frac{3+a}{a}$ s than 50%.

Also we must emphasise that even though he manages to solve item number nine $\frac{a(a-4)}{(a-4)^2}$, the item number seventeen $\frac{a(a-4)}{(4-a)^2}$ is too difficult for him.

Student on the 75th percentile rank manages to solve the first nine items which test fraction comparison, simple extension, multiplication, division and cancellation like in items $\frac{6a}{15}:\frac{3}{5}$, $\frac{x-2y}{4}\cdot\frac{8}{x-2y}$ or $\frac{a(a-4)}{(a-4)^2}$. He manages to solve partially item number ten which tests additions of fraction and numbers, but with a probability of almost 80%. He partially manages to solve items from ten to twenty-two. Among those items, there are ones that test addition like in item $\frac{5}{2a}+\frac{3+a}{a}$, sequence of calculating operations, cancellation like in item $\frac{x-2y}{8}\cdot\frac{4}{2y-x}$ and double fraction calculation. The student does not manage to solve the last two items. It is necessary to emphasise that although the probability for the student to solve the item twelve

$\frac{5}{2a}+\frac{3+a}{a}$ is almost 80%, the probability to solve item nineteen $\frac{5}{2a}-\frac{3+a}{a}$ is a little bit higher than 50%.

Student on the 90th percentile rank manages to solve the first sixteen items. These are the items which test fraction comparison, simple extension, multiplication and cancellation of fractions like in items $\frac{6a}{15}:\frac{3}{5}$ $\frac{x-2y}{4}\cdot\frac{8}{x-2y}$ or $\frac{a(a-4)}{(a-4)^2}$, addition of fraction and number, addition like in item $\frac{5}{2a}+\frac{3+a}{a}$, order of calculations, cancellation like in item $\frac{x-2y}{8}\cdot\frac{4}{2y-x}$ and calculation with double fractions.

He partially manages to solve items like: $\frac{a(a-4)}{(4-a)^2}$, $\frac{5}{2a}-\frac{3+a}{a}$ and $\frac{10x^7}{2x^3(x+1)}$. Item $\frac{10x^7}{2x^4+2x^3}$ resulted in being the most difficult one. It consists of several simple calculation steps – simple denominator factorisation, cancellation of numbers 2 and 10, and cancellation of powers x^7 and x^3. Nevertheless, the student on the 90th percentile rank manages to solve the item with the probability a little bit higher than 50%.

6.2. LINEAR EQUATIONS

Dragica Martinović, Kristina Penzar, Željka Milin Šipuš

Linear equations subject domain was tested with 24 items.

The items assessed the student's capability of:

- solving linear equations
- solving a system of two simultaneous linear equations in two unknowns
- transforming the proposed mathematical expression in algebraic form
- expressing one value in the formula by means of other values

In this subject domain the following types of items were presented:

- open type items
- simple multiple choice items
- complex multiple choice items

Table 14. Item schedule according to groups in *linear equations* subject domain

Item description	Item sequence number on graph
Solving linear equations	1, 2, 3, 4, 9, 10, 14, 15, 18, 19, 20, 21, 22, 24
Solving a system of two simultaneous linear equations in two unknowns	5, 6, 12, 13, 17
Transforming the proposed mathematical expression in algebraic form	8, 11, 16
Expressing one value in the formula by means of other values	7, 23

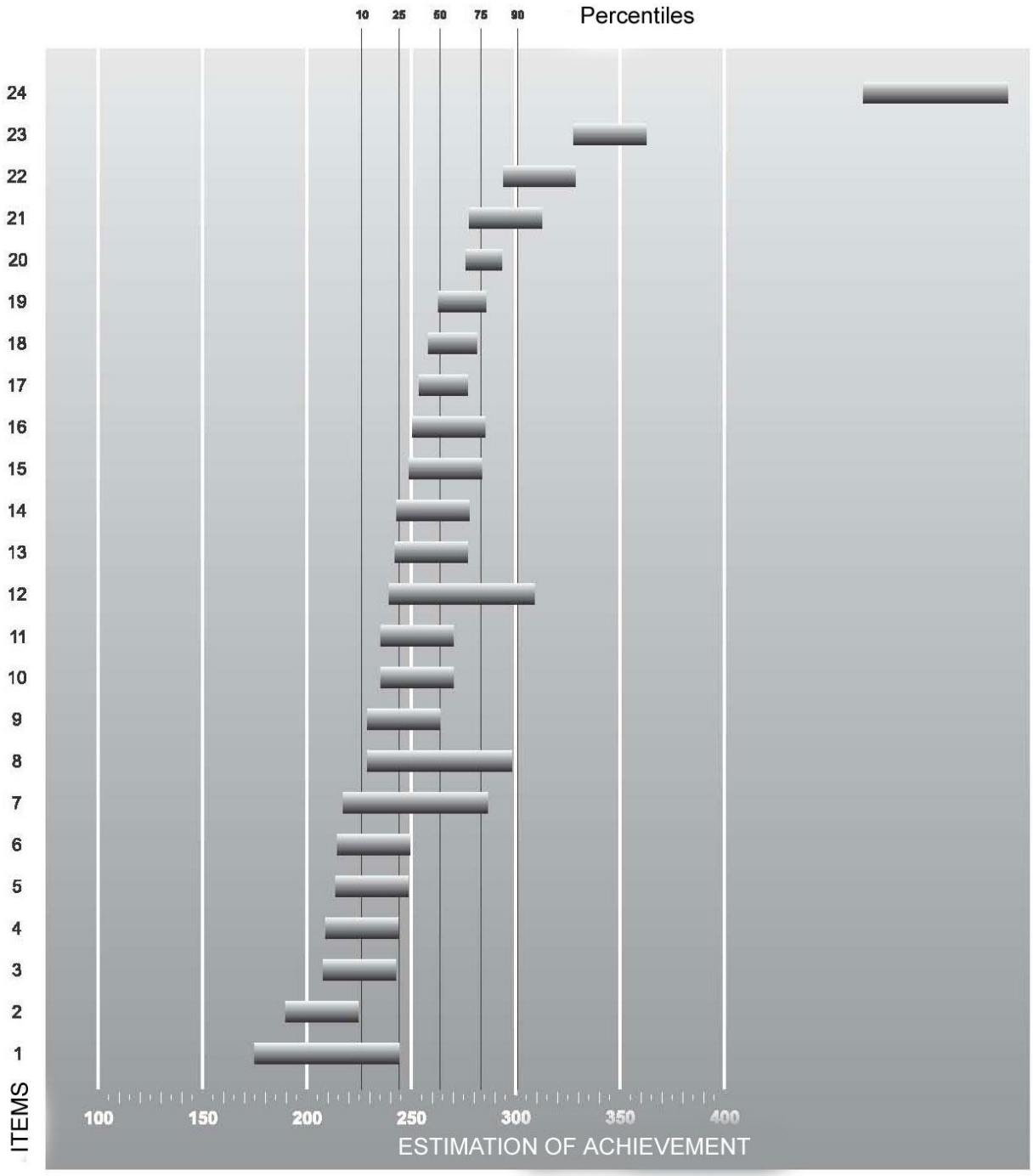
Item solving efficiency in this subject domain was between 91.5% and 1.5%. It was noted that six items was solved with efficiency higher than 80%, two items were solved with efficiency lower than 20%, and 16 items were solved with efficiency between 20% and 80%. The last item especially stands out as poorly solved item (only 1.5% of students managed to solve that item correctly).

Some types of test items, their sequence number related to ordering items by difficulty, and efficiency percentages, are listed in the following table.

Table 15. Item solving efficiency in the *linear equations* subject domain

Item sequence number on the graph	Items	Item solving efficiency
1.	Solve the equation –*x*+2=4*x*+1	84.3%
2.	Solve the equation –4*x*=9	91.5%
3.	Solve the equation system $\begin{cases} 2a+3b=5 \\ 2a-b=1. \end{cases}$	82.6%
7.	If $P=\frac{a\cdot c}{2}v$, then □ equals A. $v=\frac{2P}{a-c}$, B. $v=\frac{2P}{a+c}$, C. $v=\frac{a+c}{2P}$, D. $v=\frac{2P-a}{c}$.	70.5%
10.	Solve the equation d-(2-3d)+4=2(d-2)	71.5%
15.	Solve the equation $\frac{1}{x}+\frac{2}{x-3}=0$.	61.9%
19.	Solve the equation $\frac{x-2}{3x+1}=\frac{2x-5}{6x}$.	50.9%
20.	Solve the equation $\frac{4x-2}{3x+1}=\frac{2x-5}{6x}+1$.	39.3%
22.	Define the unknown value *x* in the equation *p*+*px*-2=0, if *p* is rational number different than 0.	28.9%
24.	Depending on *p* parameter define the solution of equation *px*-3*p*=0	1.5%

Graph 6. Item range in *linear equations* subject domain



Student on the 10th percentile rank manages to solve one item, he partially manages to solve six items, but he does not manage to solve the other seventeen items. He manages to solve the type of items like item number two from *Table 15*, therefore the simplest linear equation, which is not the easiest model item, although it is the item with the highest solving efficiency. Furthermore, the student partially manages to solve, for example, item number 5 in which it is required to solve the system of two simultaneous linear equations in two unknowns and item 7, in which it is required to express one value in the formula by means of other values.

Student on the 25th percentile rank manages to solve the first four items, he partially manages to solve the next ten items but he does not manage to solve the remaining ten items. He manages to solve items 1 and 2 from *Table 15*, therefore, the linear equations which require elementary grouping of unknowns and loose coefficients. He partially manages to solve equations and systems which involve brackets, for example, item number 10, but he does not manage to solve item number 15. He partially manages to solve the most general systems with rational solutions, although the possibility for correct solving of that kind of an item is only about 50%. Also, he partially manages to solve item 7.

Student on the 50th percentile rank manages to solve the first six items and item 9. He partially manages to solve items 7, 8, 10 – 19, but he does not manage to solve other items. Unlike other students on the 25 percentile rank, he partially manages to solve equations which transform to linear equations (items 15 and 19) and writing mathematical sentences. He does not manage to solve the item which includes solving linear equation with complex transformation (item 20).

Student on the 75th percentile rank manages to solve 15 items, he manages to solve partially nine items, but he does not manage to solve three items. He manages to solve general systems of linear equations and equations which transform into linear equation (item 15), but like the former students, he partially manages to solve item 7 in which it is required to express one value in the formula by means of other values. He partially manages to solve item which includes solving linear equation with complex transformation (items 19 and 20). He does not manage to solve item 22 and 24 which include linear equation parameter.

Student on the 90th percentile rank manages to solve 19 items, he manages to solve partially another three items, but he does not manage to solve two items. Unlike the student on 75th percentile rank, he manages to solve the item which includes solving linear equation with complex transformation (item 20). He partially manages to solve item 22, but item 24 is too difficult for him. Item 24 is exceptionally difficult; in analysis, it stands out from other items.

6.3. QUADRATIC EQUATIONS

Josipa Pavlić, Ljubica Stanković, Željka Milin Šipuš

Quadratic equations subject domain was tested with 27 items. Like the other subject domains, this one did not contain items which correlate subject matter from different subject domains (for example, applying quadratic equation in geometry). The items were open type, except for several simple multiple choice items.

The items assessed the student's capability of:

- solving quadratic equation given in the format $ax^2 + bx + c = 0$
- solving quadratic equation given in the format $ax^2 = 0$
- solving quadratic equation given in the format $ax^2 + c = 0$
- solving quadratic equation given in the format $ax^2 + bx = 0$
- solving complex quadratic equations
- among the offered equations, finding the one with given solutions, or defining the quadratic equation from the given solutions
- defining unknown parameter in quadratic equation
- using Viete's formulas

26 items scores will be presented out of 27, as one of the items has shown some deviation from the model.

Table 16. Item schedule according to groups in *quadratic equations* subject domain

Item description	Item sequence number on graph
Solving quadratic equation given in the format $ax^2 + bx + c = 0$	1, 2, 4, 7, 9, 10,19
Solving quadratic equation given in the format $ax^2 = 0$	5
Solving quadratic equation given in the format $ax^2 + c = 0$	15, 17, 20, 22
Solving quadratic equation given in the format $ax^2 + bx = 0$	13, 14
Solving complex quadratic equations	11, 12, 16, 18, 23, 24
Among the offered equations, finding the one with given solutions, or defining the quadratic equation from the given	3, 8, 21, 26

Item description	Item sequence number on graph
solutions	
Defining unknown parameter in quadratic equation	25
Using Viete's formulas	6

Item solving efficiency was between 9.1% and 81.9%.

Some types of test items, their sequence number related to ordering items by difficulty, and efficiency percentages, are listed in the following table.

Table 17. Item solving efficiency in the *quadratic equations* subject domain

Item sequence number on the graph	Items	Item solving efficiency
1.	Solve the equation $x^2 + 5x + 6 = 0$	81.9%
5.	Solve the equation $(x – 2)^2 = 0$	72.6%
8.	Which of the following equations have the solutions x = -3, x = 7 a) $x^2 – 4x -21 = 0$ b) $x^2 – 4x +21 =0$ c) $x^2 + 4x – 21 = 0$ d) $x^2 + 4x + 21 = 0$	70.2%
11.	Solve the equation $(x – 2)(x + 3)= 14$	59.1%
13.	Solve the equation $4x^2 + 20x = 0$	57.1%
15.	Solve the equation $49 = (x – 2)^2$	50.8%

Item sequence number on the graph	Items	Item solving efficiency
17.	Solve the equation $x^2 = 49$	37.3%
18.	Solve the equation $1 = \frac{7}{x^2 - 2}$	35%
25.	Define all real numbers *m* for which the equation $3x^2 - 2x + m = 0$ has the real solutions?	9.1%

Items in the group in which it was needed to solve general quadratic equation, were solved in the range from 81.9% to 66.9% if the solutions were real numbers. It is shown, for example, that the leading member sign change, or changing equation integer coefficient into a rational one, result in a slightly poorer solving of the equation. For example, the equation of this form where the coefficients are written in form of fractions was solved correctly by 74.2% of tested students. The least successfully solved equation was the one with the integer coefficients, whose solutions were irrational numbers.

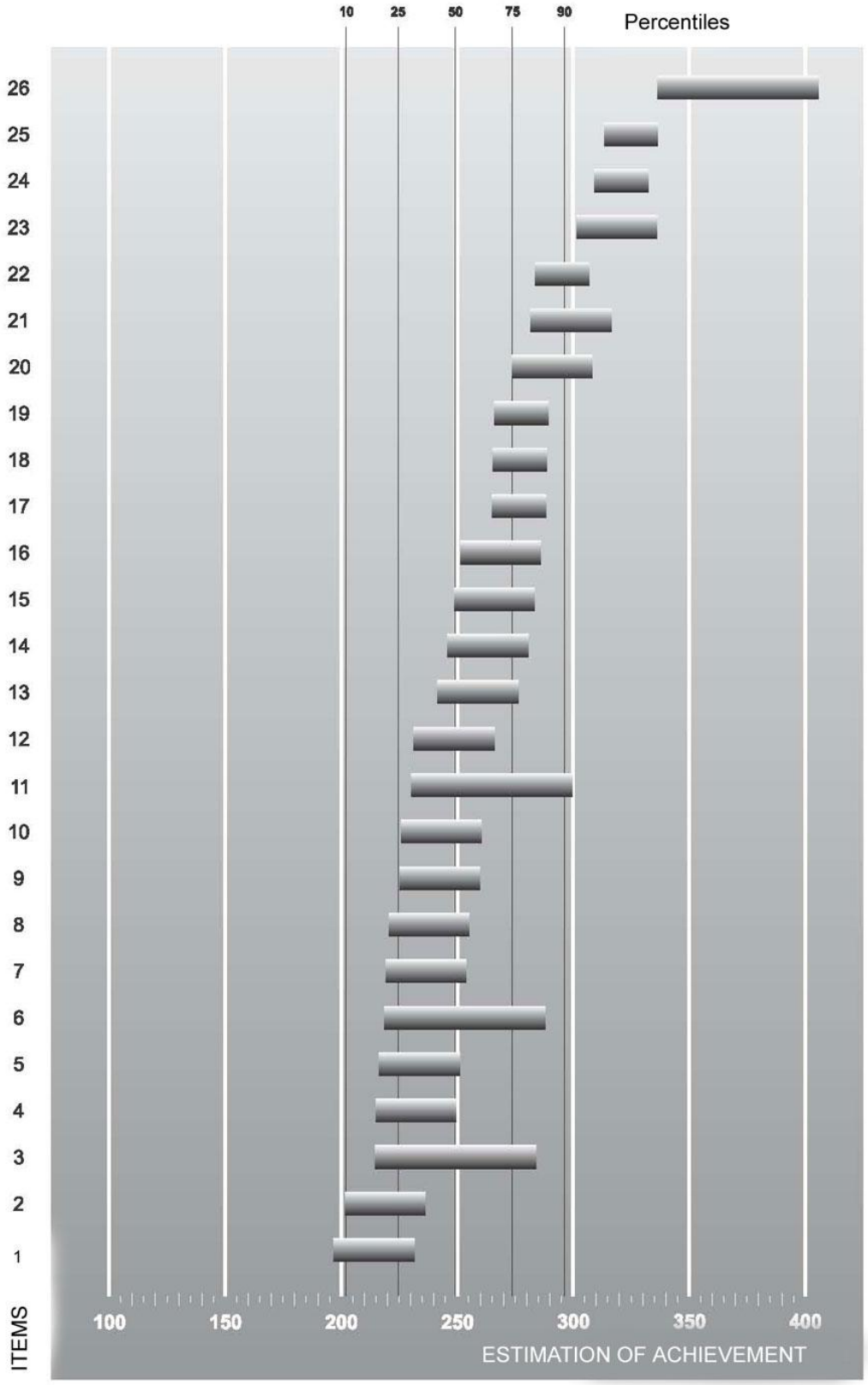
It is important to note that the general type of equation with complex number solutions was solved correctly by only 36.5% of students.

Special quadratic equations (items 5, 11, 13, 15, 17 and 18) were correctly solved ranging from 72.6% to 35%.

The lower solving percentage of items 15 and 17, as compared to item 5, should be noted. The most frequent mistake in items 15 and 17 was that the student did not see the existence of two solutions. The equation in which the unknown was not specified with *x* but with *t* was slightly less successfully solved. Furthermore, a quadratic equation analogue to the one like item 17, only with complex number solutions, was correctly solved by only 21.5% of examinees.

The complex quadratic equations were solved ranging from 62.3% to 16.9%, while biquadrate equation was correctly solved by 10.7% of examinees.

Graph 8. Item range in *quadratic equations* subject domain



Student on the 10th percentile rank does not manage to solve any of the items from this subject domain. He manages to solve the easiest item in this domain only partially, item 1, but the very first next item from this domain is too difficult for him.

Student on the 25th percentile rank manages to partially solve items 1 – 8, while the other items are too difficult for him. The probability of solving items 3 – 8 is slightly higher than 50%. That student finds too difficult all of the equations which have fractions for coefficients, which have non integer solutions and also special and complex quadratic equations.

Average student, i.e. student on the 50th percentile rank manages to solve items 1 and 2, and almost manages to solve items 4 and 5, which means that he successfully manages to solve the simplest and common quadratic equations with rational coefficients whose solutions are integers. As soon as the coefficients are big numbers, or numbers with different signs, or a certain coefficient is equal to zero, this student manages those items partially (items 3, 6 – 14). He does not manage to solve the remaining 12 items. That means that there is a small possibility for the student on the 50th percentile rank to solve some of the complex quadratic equations or equations with complex solutions. That student will correctly solve equation $x^2 = 49$ with less than 50% possibility.

Student on the 75th percentile rank manages to solve 9 items (items 1, 2, 4, 5, 7 - 10, and 12), he manages to solve partially ten items (items 3, 6, 11, 13 – 19), but he does not manage to solve seven items. That means that the student manages to successfully solve integer or rational coefficients quadratic equations, but poorly manages to solve special and complex quadratic equations. He partially manages to solve the quadratic equation $x^2 = 49$.

Student on the 90th percentile rank i.e. student who belongs in the top 10% best students, manages to solve 18 items included in this subject domain (first nineteen items, except item 11), partially manages to solve four items, and does not manage to solve the last 4 items. He manages to solve the quadratic equation$x^2 = 49$, but he manages to partially solve the analogue equation which includes irrational solutions. He manages to solve the general quadratic equation with complex solutions, and he partially manages to solve the equation analogue to the $x^2 = 49$ equation but with complex solutions.

Like the discussion about existence of the real solutions to quadratic equation, the biquadrate and complex quadratic equations are too difficult for the student on the 90th percentile rank.

6.4. QUADRATIC FUNCTIONS

Jagoda Krajina, Dobrila Golubović, Željka Milin Šipuš

Quadratic functions subject domain was tested with 24 items.

In this subject domain, the following types of items were presented:

- open type items
- simple multiple choice items
- complex multiple choice items

The items assessed the student's capability of:

- calculating the value of function, and vice versa, defining the argument value from the known quadratic function value
- defining quadratic function zeros, and vice versa, defining the quadratic function from the known zeros,
- defining the function vertex, i.e. the quadratic function minimum and maximum
- identifying and defining the quadratic function graph
- interpreting the effect of leading coefficient on the function graph
- solving a quadratic inequality

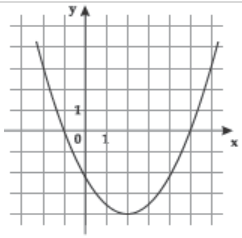
Table 18. Item schedule according to groups in *quadratic functions* subject domain

Item description	Item sequence number on graph
Calculating the value of function, and vice versa, defining the argument value from known quadratic function value	1, 2, 3, 4, 10
Defining quadratic function zeros, and vice versa, defining quadratic function from the known zeros	5, 11, 12, 13, 17, 21

Item description	Item sequence number on graph
Defining the function vertex, i.e. quadratic function minimum and maximum	23
Identifying and defining quadratic function graph	6, 7, 9, 14, 15, 19, 20, 22
Interpreting the effect of leading coefficient to function graph	8
Solving a quadratic inequality	9, 14, 16, 18, 24

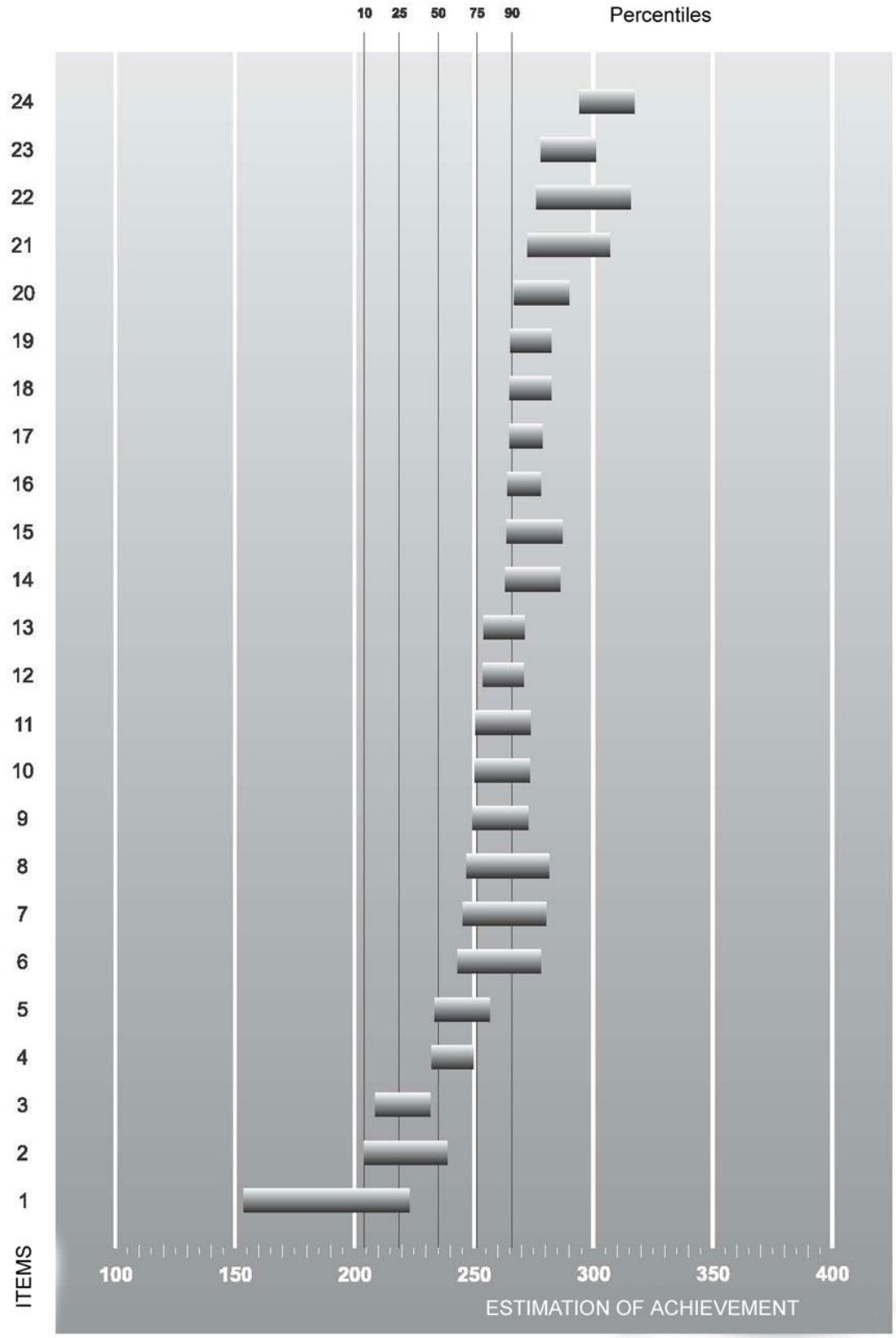
Item solving efficiency in this subject area was ranging between 7.8% and 81%. Some types of test items, their sequence number related to ordering items by difficulty, and efficiency percentages, are listed in the following table.

Table 19. Item solving efficiency in the *quadratic functions* subject domain

Item sequence number on the graph	Items	Item solving efficiency
1.	For function f(x) = x^2 define positive real number x if f(x) = 9.	81.0%
2.	For function f(x) = x^2 define f(-2).	79.6%
3.	For function f(x) = x^2 define negative real number x if f(x) = 9.	70.4%
5..	Define the zeroes of the function f(x) = x^2 +x – 2.	55.8%
9.	f(x) function graph is presented in the picture. Write the interval for which it is f(x) > 0 	32.4%
16.	Solve the inequality $x^2 + x – 2 \geq 0$	17.8%
19.	In coordinate system build the graph of the function f(x) = - x^2 + x + 2. It is necessary to sketch the zeroes and the function vertex.	19.3%
23.	Define the minimum value of the function f(x) = $(x\text{-}1)^2$ + 3.	11.7%
24.	Solve the inequality $x^2 + x + 2<0$.	7.8%

It can be noted that the students had better results in solving quadratic equations in the subject domain of *quadratic equations*, than they managed to define quadratic function zeroes in the *quadratic functions* subject domain.

Graph 9. Item range in *quadratic functions* subject domain



Student on the 10th percentile rank manages to partially solve one item, while he does not manage to solve the remaining twenty-three items. He partially manages to solve the first two items from *Table 19*. With a probability of about 70%, he can define the positive number whose square equals 9, while he defines the value of the function f(x) = x^2 in x = – 2 with the possibility of 50%.

Student on the 25th percentile rank manages to partially solve three items while he does not manage to solve the remaining twenty one items. Unlike the student on the 10th percentile rank, he partially manages to solve the item in which it is necessary to define the negative number whose square equals 9. All the other variations of those items (defining the value of the complex quadratic function, that is, the argument for the assigned value of complex quadratic function) are too difficult for him.

Student on the 50th percentile rank manages to solve the first and the third item. He partially manages to solve items two, four and five, while he does not manage to solve the other items. Therefore he manages to solve the items in which positive and negative argument x must be defined for the given value of the function f(x) = x^2, while he partially manages to solve items in which it is needed to define function value at the point, that is, the general quadratic function zero. He correctly defines the zero of the general quadratic function with a probability of about 50%, while he does not manage to solve the items in which special quadratic function zeroes must be defined. He needs a designation of function in factorized form, and uses the concept of zeroes for defining the function coefficients. He does not manage to solve items in which it is required to define and sketch quadratic function graph.

Student on the 75th percentile rank manages to solve the first four items, he manages to solve partially the next seven items, but he does not manage to solve the remaining thirteen items. In simple situations, that student can calculate the value of the function and vice versa, define the value of the argument from the given value of the quadratic function. He partially understands the concept of zero of the general and special complex function, but he does not manage the item in which it is necessary to factorize quadratic function. He partially manages to solve items in which it is necessary to identify quadratic function graph, to interpret the effect of the leading coefficient on the function graph, and to use the term of function maximum (minimum). He partially manages to solve the item in which it is required to graphically solve the quadratic inequality, but the probability of solving that item correctly is somewhat higher than 50% (item 9). He does not manage to solve the items in which it is required to solve the quadratic inequality (item 16 type), and also the items in which it is required to sketch the quadratic function (item 19 type).

Student on the 90th percentile rank manages to solve the first five items; he manages to solve partially fifteen items, while he does not manage to solve four items. Unlike the student on the 75th percentile rank, he also partially manages to solve items 14 – 20, therefore, items in which it is required to identify or sketch a quadratic function graph, to interpret the effect of leading coefficient on function graph, to define the function minimum or maximum, and to solve quadratic inequality (items like 16 and 19).

It should be noted that four items are still too difficult for the Student on the 90th percentile rank. In the penultimate item, it is required to define the minimal function value. Many of the students did not even try

to solve that item, but among the ones that tried to solve it, a large part of them gave the x instead of *f(x)* function value as the answer.

The most difficult item in the subject domain quadratic inequality is the one which has no solution (item 24 type). That item was solved only by 7.8% students. Since the correspondent quadratic equation has no real solutions (correspondent function graph is above x-axis in full), many of the students defined its complex solutions and also offered for the inequality solution set $\left\langle -\frac{1}{2}-i\sqrt{7}, -\frac{1}{2}+i\sqrt{7} \right\rangle$.

6.5. TRIANGLE

Miroslav Smuđ, Ljubica Stanković, Željka Milin Šipuš

Triangle subject domain was tested with 27 items. Most of the items were open type, and several were simple multiple choice closed type items.

The items assessed the student's capability of:

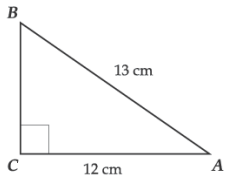
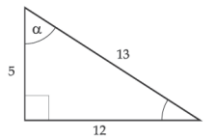
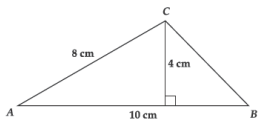
- defining triangle elements, characteristic points, using angle measures and triangle side lengths ratios, and defining angle measures on an elementary level
- identifying similar triangles and using their properties
- using Pythagorean theorem
- using right-angled triangle trigonometry
- calculating perimeter and area of triangle, including Heron's formula

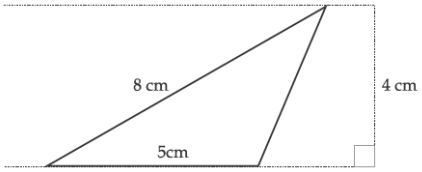
Table 20. Item schedule according to groups in *triangle* subject domain

Item description	Item sequence number on graph
Defining triangle elements, characteristic points, using angle measures and triangle side lengths ratios, and defining angle measures on an elementary level	2, 5, 6, 9, 10, 14, 22
Identifying similar triangles and using their properties	3, 19, 26
Using Pythagorean theorem	1, 4, 17, 24
Using right-angled triangle trigonometry	11, 12, 13, 15, 16, 20, 21
Calculating perimeter and area of triangle, including Heron's formula	7, 8, 17, 18, 23, 24, 25, 27

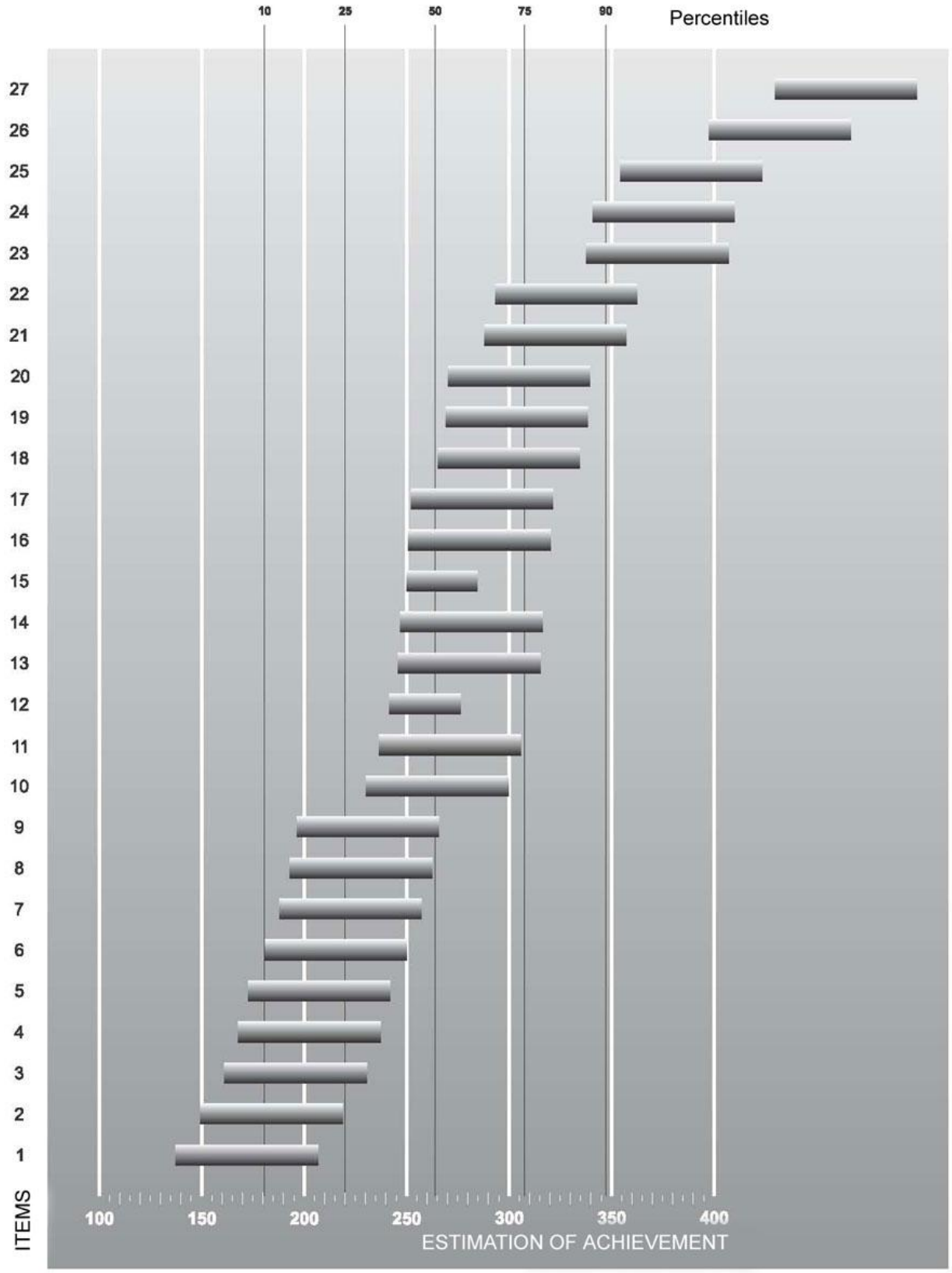
During the testing, item solving efficiency was between 88.1% and 7.1%. Some types of test items, their sequence number related to ordering items by difficulty, and efficiency percentages, are listed in the following table.

Table 21. Item solving efficiency in the *triangle* subject domain

Item sequence number on the graph	Items	Item solving efficiency
1.	Define the unknown side of right-angled triangle from the picture. 	88,1%
2.	The degree of the right-angled triangle acute angle is 47°. What is the degree of that triangle's other acute angle?	85.9%
4.	What is the length of the right-angle triangle hypotenuse with the legs 3 cm and 8 cm long	82.1%
7.	Calculate the triangle area that has one side 7 cm long, and the height to that side is 6 cm in length.	76.2%
11.	In a right-angle triangle the lenghts of the sides are equal to a= 5 cm, b= 12 cm, c= 13 cm. Find sin α.	60.3%
12.	Define sin α in the right-angle triangle in the picture. 	58.9%
18.	Define the area of the ABC triangle in the picture. 	48.5%

Item sequence number on the graph	Items	Item solving efficiency
21.	Define the right-angle triangle leg length adjacent to angle α= 40° if the length of the height to the hypotenuse of that triangle is 4 cm.	40.9%
25	Define the area of the ABC triangle in the picture. 	20.6%
27.	The given triangle has sides of 4 cm, 5 cm and 6 cm. What is the area of that triangle?	7.1%

Graph 10. Item range in *triangles* subject domain



Student on the 10th percentile rank does not manage to solve any of the items, and he partially manages to solve the easiest five items. That means that he partially manages to solve the Pythagorean theorem (items 1 and 4 in Table 21) and to define the angle degrees in a special (right-angle) triangle (item 2 fin Table 21). He does not manage to solve the rest of the items.

Student on the 25th percentile rank manages to solve the two easiest items, partially manages to solve next seven items, but does not manage to solve the rest of the 18 items. That means that he manages to solve the Pythagorean Theorem, while he partially manages to define the triangle angle degree on an elementary level (using the sum of the triangle angle measures), and calculating the triangle area using the formula $P = \frac{av}{2}$ (item 7) and the right-angle triangle area.

Student on the 50th percentile rank manages to solve eight items; he partially manages to solve nine items, but does not manage to solve 10 items. That means that he manages to solve items which the student on the 25th percentile rank partially manages to solve (elementary defining of the triangle angle degree, calculating the triangle area using formula $P = \frac{av}{2}$), and he partially manages to solve the ratio between the degree of angles and side lengths, and also the right-angle triangle trigonometry, regardless of the fact whether the triangle is defined with a sketch or data (items 11 and 12). He does not manage to solve item 18 in which it is required to calculate the triangle area defined with more elements than necessary.

Student on the 75th percentile rank manages to solve thirteen items; he manages to solve partially another nine items, but he does not manage to solve the remaining five items. He manages to solve the items which test right-angle triangle trigonometry (item 11 and 12), having a bigger possibility of solving correctly that kind of item if the triangle is given in a sketch. Unlike the 50th percentile student, he partially manages to solve item 18 and item 21 of trigonometry application, using more complex analysis.

Student on the 90th percentile rank manages to solve twenty items, he manage to solve four items partially, while he does not manage to solve three items. He manages to solve item 18, but he does not manage to solve item 25 in which it is required to calculate the obtuse-angled triangle given by formula $P = \frac{av}{2}$, but in the item there is a need for careful interpretation of the triangle basis and the height dropped on it. For the 90th percentile student the Heron's formula for the triangle area determination is too difficult.

6.6. MATHEMATICAL LITERACY

Miroslav Smuđ, Ljubica Stanković, Željka Milin Šipuš

Mathematical literacy subject domain was tested with 24 items. Most of the items were open type, and several were simple multiple choice closed type items.

Mathematical literacy domain is highlighted as a special domain for determining the competence of applying mathematical ideas in real life situations. The development of that capability is one of the key objectives in mathematical education.

A very important part in solving these items in this domain is the mathematisation of the problems (formulating the problems in mathematical language terms). Furthermore, the items assessed the student's capability of:

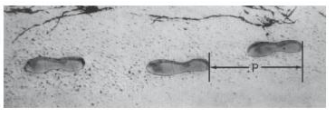
- connecting and calculating on an elementary level, and using simple algebraic correlations
- using the percentage calculus
- solving linear equation and equation systems
- calculating lengths, perimeter and area on an elementary level, and converting units
- using Pythagorean theorem, similarity and trigonometry in defining measureable object characteristics

Table 22. Item schedule according to groups in *mathematical competence* subject domain

Item description	Item sequence number on graph
connecting and calculating on an elementary level, and using simple algebraic correlations	1, 3, 5, 6, 7, 20, 21
using the percentage calculus	9, 10, 11, 13, 14, 15, 16
solving linear equation and equation systems	2, 12, 17
calculating lengths, perimeter and area on an elementary level, and converting units	4, 12, 21
using Pythagorean theorem, similarity and trigonometry in defining measureable object characteristics	8, 18, 19, 22, 23, 24

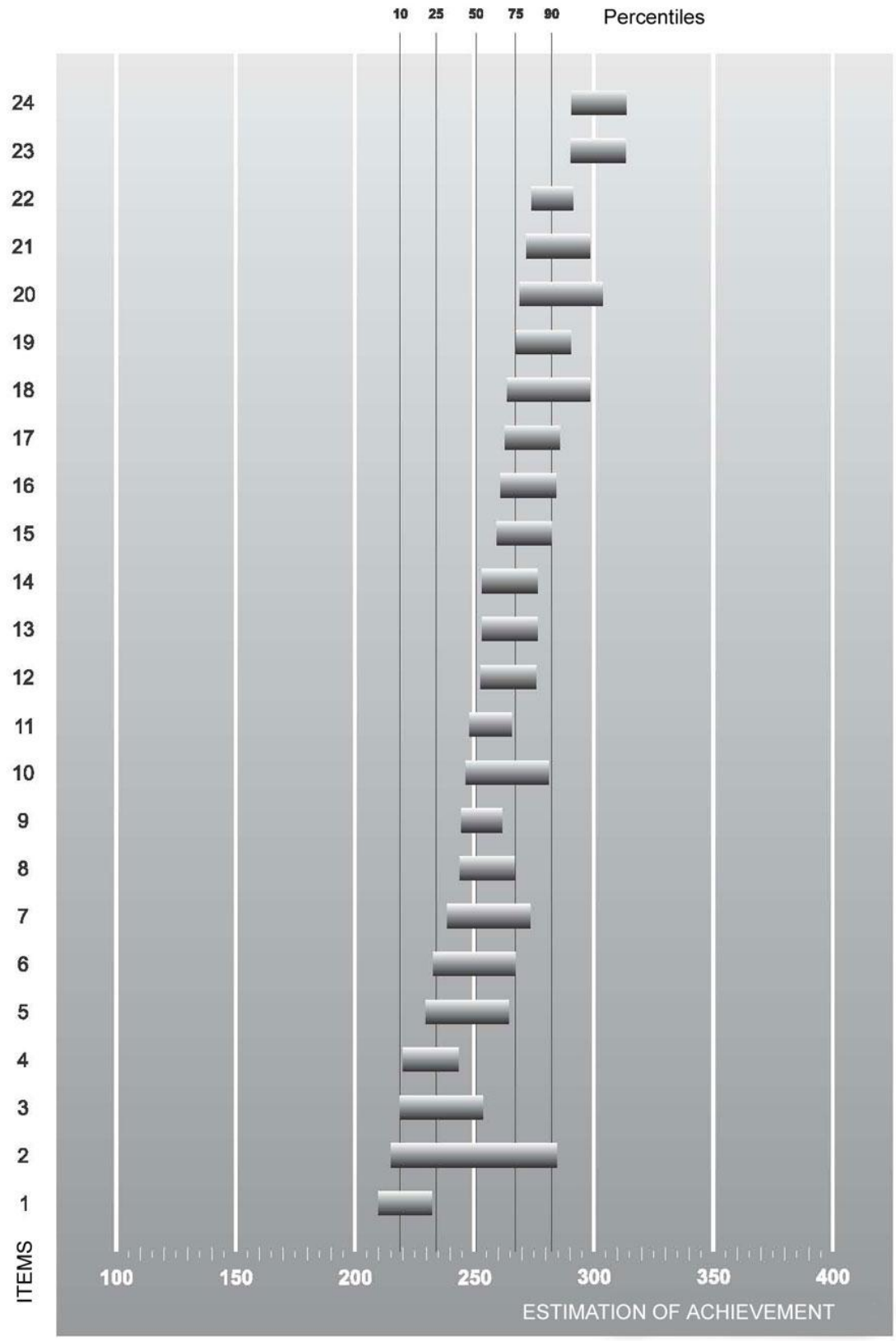
During the testing item solving efficiency was between 14.3% and 86.4%. Some types of test items, their sequence number related to ordering items by difficulty, and efficiency percentages, are listed in the following table.

Table 23. Item solving efficiency in the *mathematical literacy* subject domain

Item sequence number on the graph	Items	Item solving percentage
1.	The picture shows a man's footprints. The length of the step *P* is defined as thedistance between the ends of the two adjacent footprints.  The connection between *n* and *P* is approximately given for the man with the formula $\frac{n}{P}=140$, where: *n* = number of steps per minute *P* = step length If the formula is valid also for Peter who walks 70 steps per minute, what is the length of his step? Show how you got the results.	86.4%
2.	In the basket there are 89 balls – some of them are small, and some are big. Every small ball weighs 2 g, and every big ball weighs 5 g. The total ball weight in the basket is 256 g. How many of the small balls are in the basket? A.115 B.63 C.26 D.25.6	65.8%
4.	A square plate has the area of 64 cm^2. Define the dimensions of the plate.	78.3%
8.	A 7-metre-long beam is leaning against a building. If the bottom of the beam is moved away 2 m from the building base, what hight will the beam reach?	57.8%
9.	In the Adriatic sea there are 698 islands. Only 23 of them	58.1%

Item sequence number on the graph	Items	Item solving percentage
	are inhabited. What percentage of the island is inhabited?	
10.	Ana is *a* cm tall, and Branka is *b* cm tall. The expression $b = a + 0.15\,a$ describes that: A. Branka is 0.15 cm taller than Ana B. Branka is 15% taller than Ana C. Ana is 15 cm taller than Branka D. Ana is 0.15% taller than Branka	55.3%
12.	A rectangular plate has the area of 192 cm^2. The length of the plate is three times its width. Define the length of the plate.	47.9%
13.	There are 698 islands in the Adriatic sea. Only 23 islands are inhabited. What percentage of the islands is **not** inhabited?	46.6%
19.	At what angle to the ground should a 3.5-metre-long ladder be leaning in order to reach 2.3 meters in height?	32.5%
21.	Ivan knows that the length of his step is 0.80 metres. For Ivan's walking, the formula $\frac{n}{P} = 140$ applies, where: *n* = number of the steps per minute *P* = length of the steps. Define Ivan's speed in metres per second, and in kilometres per hour. Show how you reached that result.	35.4%
24.	A box is 12 cm wide and long, and 6 cm high. How long is the longest stick that can fit into the box?	14.3%

Graph 11 Item range in *mathematical literacy* subject domain



Student on the 10th percentile rank does not manage to solve any of the items, and he partially manages to solve three items. He partially manages to solve the simplest item 1 and item 2. Item 2 (linear equation system) was solved by 65.8% of students, but all students in the sample have difficulties solving with that kind of an item. That kind of items is not adequately solved by the best student sample.

Student on the 25th percentile rank manages to solve only the easiest item 1; he partially manages to solve five items ,but he does not manage to solve the remaining 18 items. Among the items which the student does not manage to solve are all items with the percentage calculus and complex geometric measuring items. For example, student on the 25th percentile rank partially manages to solve items 2 and 4, and he does not manage to solve the easiest item, in which percentage calculus is used (item 9), and the item which examines the usage of Pythagorean Theorem.

Average student, i.e. student on the 50th percentile rank manages to solve two items; he partially manages to solve nine items but he does not manage to solve the remaining 13 items. For example, he manages to solve item 4, partially manages to solve items 8, 9 and 10, but he does not manage to solve items 12 and 13. Item 12 is analogue to item 4 (plate is rectangular instead of square shape, and item 13 is analogue to item 9 (the question is put negatively: *What percentage of the islands is not inhabited?).*

Student on the 75th percentile rank manages to solve 8 items; he partially manages to solve 11 items, and he does not manage to solve 5 items. For example, he manages to solve item 8 (usage of Pythagorean Theorem) and item 9 (the simplest item testing percentage calculus), he partially manages to solve items involving percentages, including item 10, and does not manage to solve items 19 (trigonometry application), and 21.

Student on the 90th percentile rank, that is, the student ranking in the top 10% of the students, manages to solve 14 items, partially solve 8 items, and does not solve 2 items. For example, he solves all of the items that involve percentages, including item 10, which indicates a sound understanding of the concept of percentage; he solves items involving simple geometry measures (item 12) and simple application of the Pythagorean Theorem (item 8). He partially manages to solve the item involving trigonometry application (item 19), and item 21. However, it should be noted that item 24 (applying Pythagorean Theorem to the spatial diagonal of a quadrant) is too difficult even for this student.

7. BACKGROUND QUESTIONNAIRES

Maja Reberšak, Biljana Vranković

In order to better understand an educational system, including Croatian, we need to look at it in its contextual frame. Students are active parts of the teaching process and together with other participants they form the basis for the development and welfare of society. In this sense, students' achievements are to be considered regarding the environment they live in, they are educated in and they grow up in. In order to obtain the best possible description of the contextual frame of the students who are test-takers, upon finishing the test booklets they were given especially designed questionnaires.

7.1. PURPOSE AND STRUCTURE OF BACKGROUND QUESTIONNAIRES

The purpose of the background questionnaires was to retrospectively obtain information on the socio-economical background of the students, as well as their attitudes and opinions of the subjects tested.

The background questionnaire administered retrospectively to students consists of two parts: the first part is general with the purpose of obtaining information on students' sex, age and socio-economical background, whereas the second part is connected to the subject being tested. The general part was designed by the group for the background questionnaires, and the questions for the second part connected to the subjects being tested were designed with the help of the groups for Croatian language and Maths.

The first questionnaire, called PU-1, was administered to students taking the test in Croatian language and it consists of two parts:

1. general part
2. Croatian language.

The second questionnaire, called PU-2, was administered to students taking the test in Maths and it consists of two parts:

1. general part
2. Maths.

The general part was the same for all students and it consists of 17 questions.

The aim of the second part of the questionnaire was to yield information on students' attitudes and opinions of the tested subjects. The questionnaire for Croatian language (PU-1) consists of 22 questions and it refers to students' opinion of the test, to their reading, to their motivation to learn the Croatian language, to their will to apply and teach what they have learnt. The questionnaire for Maths (PU-2) consists of 24 questions and it refers to students' opinion of the test, to their attitudes toward the subject tested and their motivation, to the teachers, to extracurricular activities, to the teaching methods and types of assignment used and the possible correlation to other subjects as well as to the teaching materials and aids.

Both background questionnaires were pre-tested using the cognitive labs method on sample students. Considering the analysis of the obtained information both questionnaires have been altered and designed for the main test administration.

7.2. COGNITIVE LABS

The cognitive labs method has been used to determine that the students comprehend the questions as assumed by the test designers and it is aimed at testing the questionnaire itself rather than the main test administration. It studies the mental processes one uses when completing an item. Therefore, the sample subjects (students) have to be carefully selected. Six second grade gymnasium high school students were chosen - three female and three male students. The students were different in terms of their achievements, extracurricular activities, socioeconomic status and the environment they came from.

Verbal Probing Technique was used for the cognitive lab administration. The interviewer asks the target question and the subject answers it, but the interviewer then follows up by probing for other specific information relevant to the question or the specific answer given. This gave effective insights into the functioning of the students mind. Each subject (student) spent an hour with three female testers giving answers and discussing asked questions. Having obtained all the responses and after an analysis, the group decided that some questions had to be altered whereas some were omitted. These altered questions were then used in designing the final background questionnaires which were pre-tested to ensure their validity for the main test administration. The response interpretation refers to the results obtained on the main test administration.

7.3. RESPONSE INTERPPRETATION

All students who took the tests in Croatian language and in Maths were administered the background questionnaire. The general part was the same in both questionnaires, PU-1 and PU-2, so the responses were analysed together. Of all students who took the tests 2.8% of them did not fill out the general part. 2.3% of the students who took the test in Croatian language did not fill out the background questionnaire, whereas 2.2% of the students who took the test in Maths did not fill out the background questionnaire.

All response interpretation in the following three subchapters refer to the sample of second grade gymnasium high school students.

7.3.1. General part of the background questionnaire

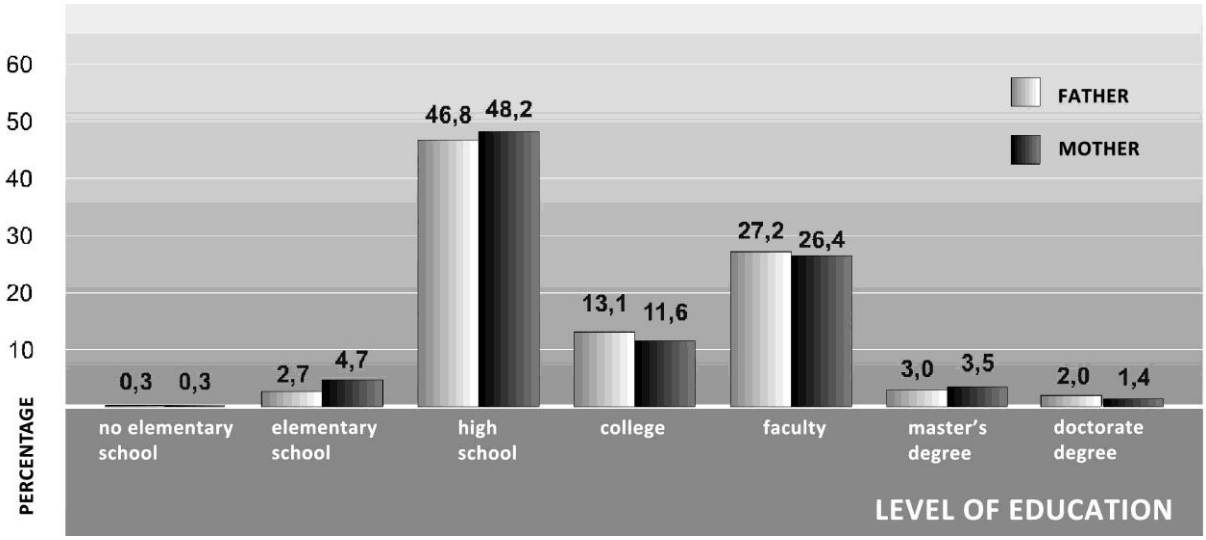
The general part of the background questionnaire was designed to yield information on sample students' demographic and socioeconomic features. The response interpretation of the general part of the background questionnaire raises several issues concerning both mentioned categories and the information about the sample students' responses.

Demographic features of the sample students. 2230 students took part in the testing, out of which 1252 were female and 978 male students. Majority of sample students, over 70% of them, were seventeen-year-old, 26,1 % were sixteen-year-olds and 1% were eighteen-year-olds.

Socioeconomic status of the sample students. The information on socioeconomic status of the sample students was yielded with *questions on their parents' level of education and employment status as well as on financial status*. Questions aimed at yielding information on their family's financial status referred to their pocket money, whether they had their own room, whether they owned a holiday house, the number of cars and television sets they had.

Students were asked questions on their parents' level on education[18]. *Graph 12* shows that the greatest percentage of parents finished high school. Next, almost one third of fathers and one third of mothers have an academic degree. A bit over 10% of fathers and mothers graduated from college. Finally, the data shows that 5% of parents have a master's degree or a doctorate degree.

Graph 12. Level of education of sample students' parents



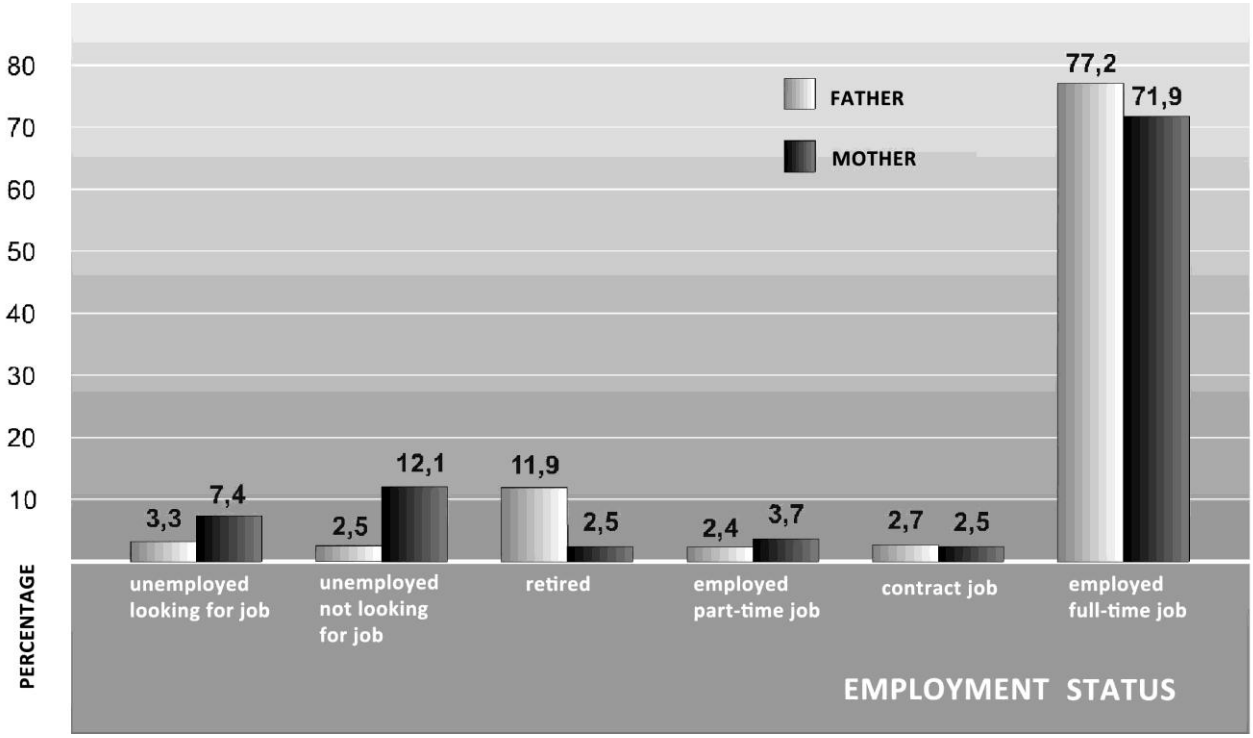
As depicted by *Graph 13*, majority of sample students' parents are employed full time. Other categories have a significantly lower percentage. Majority of sample students' parents are employed full time. Other categories have a significantly lower percentage.

The summative indicator of the financial status based on the students' responses in the general part of the questionnaire clearly shows the financial situation of the sample students' family (*Graph 14*).

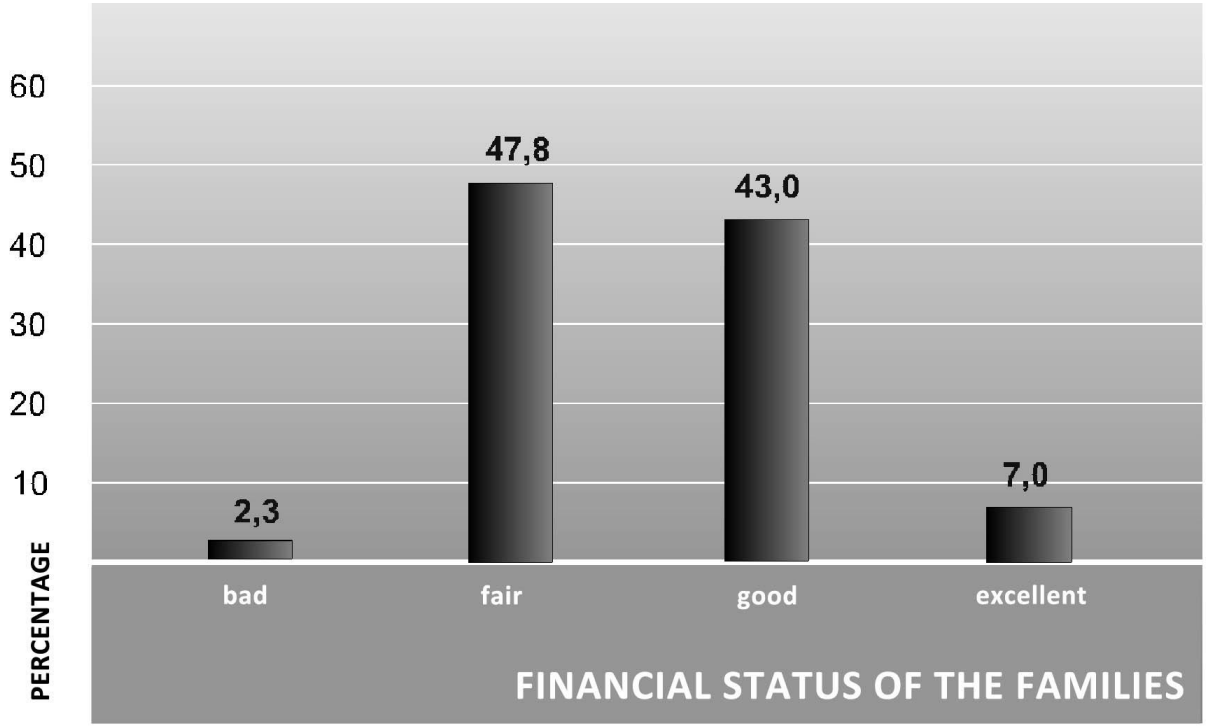
According to these results, regarding the financial status the sample students belong to one of the four groups: bad, middle, good or wealthy financial status. The distribution of these socioeconomic groups is represented in the Graph 14. The significantly big group of sample students (86,2%) belongs to the two middle groups.

18 Questions clearly stated that questions referring to students' parents referred to foster parents as well.

Graph 13. Employment status of the sample students' parents



Graph 14. Financial status of the sample students' families

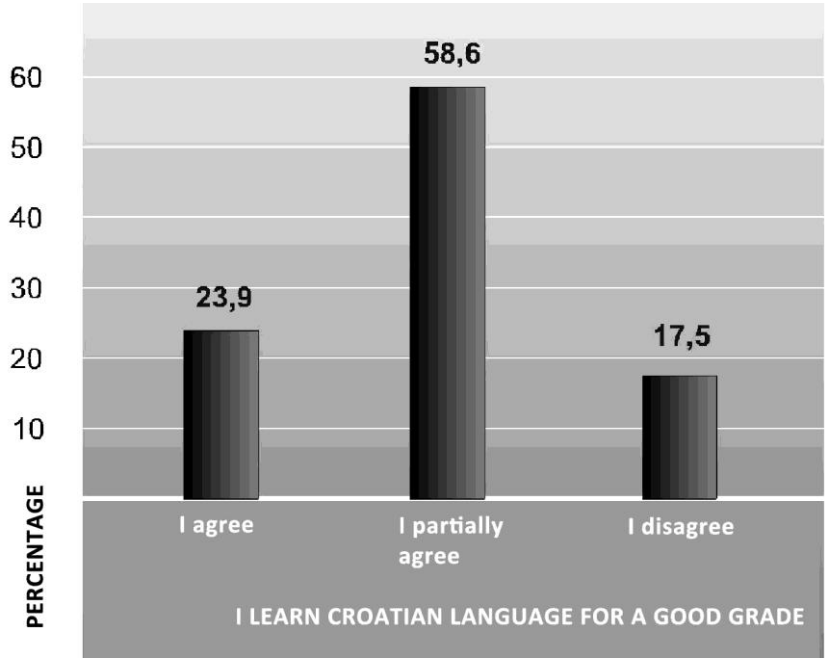


7.3.2. Croatian language

The part of the questionnaire for Croatian language (PU-1) consists of 22 questions and it refers to curricular and extracurricular experiences of the students connected to Croatian language. This part of the *Report* deals with the results of the background questionnaire connected to Croatian language. The response interpretation refers to the students attitudes toward some aspects of *motivation to learn*, *experiences in reading* and *lessons of Croatian language.*

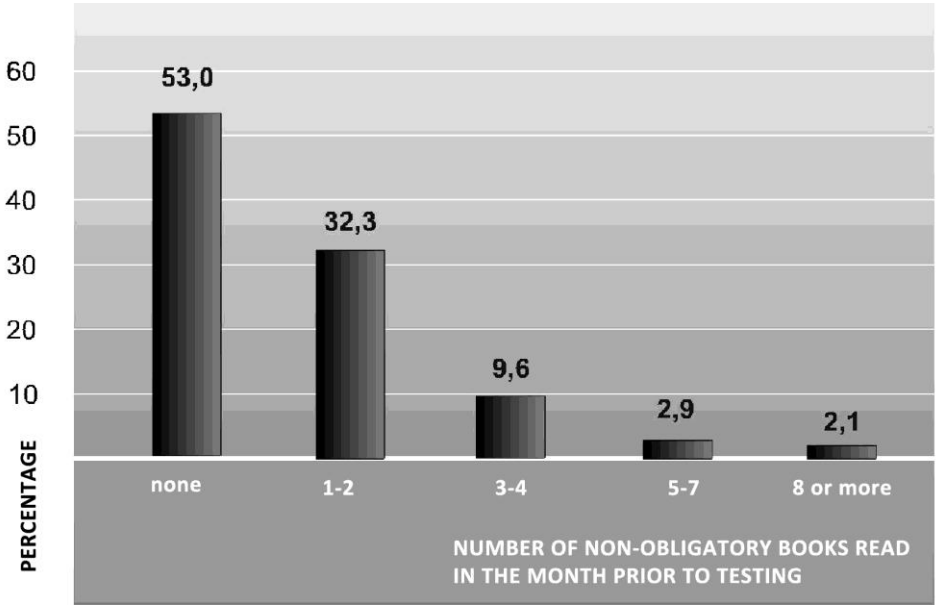
Motivation. The questions about the external motivation was aimed at yielding information on students' motivation to learn Croatian language, i.e. to what extent they learned Croatian language for a good grade. *Graph 15* shows that more than 50% of students thought that the grade was not the only reason they learned Croatian language. Less than 40% of students considered the grade the only reason to learn Croatian language, whereas less than one fifth of the students did not agree with the statement that they learned Croatian language because of the grade.

Graph 15. Motivation to learn Croatian language



Reading experience. One of the questions referred to gathering information on their reading habits and their attitudes toward reading. It was indeed interesting to see how much books, that were not obligatory, the students read last month. *Graph 16* shows that almost a third of the students did not read a book apart from the ones assigned. Half of the sample students read one or two books in the month prior to the testing, whereas a significantly low number of students reported having read more than 2 books.

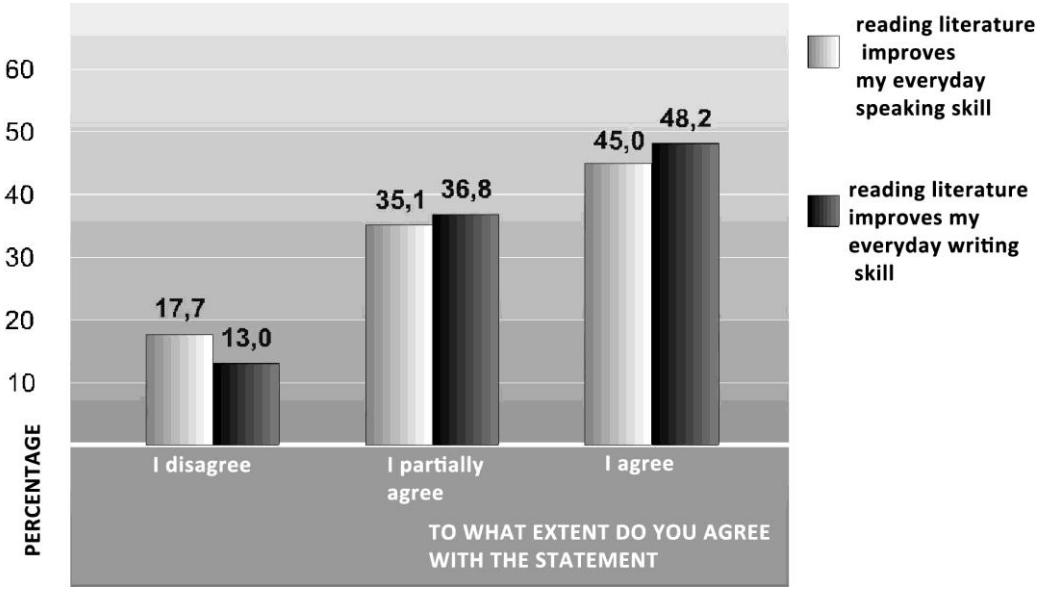
Graph 16. Number of books the students read apart from the reading assignments in the month prior to the testing



The sample students were also asked questions about how interesting the assigned readings were. Majority of students (58,3%) thought that some books were interesting whereas 23.9% of the students considered all of the books boring. 13,4 % of them found a few interesting books on the reading assignment list, and 2% of the students thought of all the books on the list were completely uninteresting.

This part of the questionnaire referred to whether the students thought that their reading habits influenced their everyday speaking and writing. As depicted by *Graph 17*, majority of students agreed with the statement that reading literary texts helped them speak and write better in their everyday life. About a third of students saw a partial influence of reading literary texts on their everyday speaking and writing skills. Less than one fifth of the sample students disagreed with the statement.

Graph 17. Students' opinion on the influence of reading on their other linguistic abilities



Croatian language classes. Questions connected to classes of Croatian language refer to teaching methods, individual lesson phases and individual unit contents. Some of the sample students' responses connected to Croatian language lessons are given in *Table 24*.

Table 24. Experiences in Croatian language classes.

	Never	***Sometimes***	***Often***
We have different forms of writing in Croatian language lessons.	9.5 %	61.7 %	26.6 %
We learn grammar in Croatian language lessons	9.5 %	57.6 %	37.3 %
We repeat lesson contents of writing in Croatian language lessons.	11.2 %	53.5 %	33.1 %
We work in pairs and groups in Croatian language lessons.	41.4 %	47.1 %	9.1 %
Croatian language lessons take place outside the classroom.	58.5 %	37.3 %	2 %

Almost half the students said that they were sometimes encouraged to learn Croatian by exploring. 16.3% students reported being very often encouraged to learn new contents by exploring, whereas one third of the students said they were never encouraged to learn Croatian by exploring.

Teaching techniques have further being explored with questions on pair and group work. *Table 24* shows that almost half the sample students sometimes works in pairs, but approximately the same number has never worked in pairs. The remaining number of the students reported frequent work in pairs and groups. Over 50% of the students said they have never had Croatian language lessons outside the classroom. About 30% of students said they quite often had this type of Croatian language lessons, whereas a small number of students reported having lessons outside the classroom very often.

Different types of Croatian language lessons have been explored with two questions. The first questions referred to the extent to which they repeated the contents during the lessons. *Table 24* clearly shows that about 50% of the students sometimes repeated the contents during the lessons, one third reported repeating the contents often, and more than one tenth of the students reported never repeating the contents during the Croatian language lessons.

The second question connected to the different types of Croatian language lessons referred to **feedback students get during classes**. 29.8% of students reported they always or often got feedback on their performance during the classes, half the students reported this to occur sometimes, whereas 17% of students said they never got feedback on their performance during the classes.

How students felt about the contents of Croatian language lessons was explored with two questions. *Table 24* clearly shows that much more than 50% of the students sometimes wrote some form of writing, about one fourth of the students wrote often, a only about one tenth of the students reported they never wrote during Croatian language classes. More than 50% of the students said they sometimes learn grammar, more than one third of the students said they learned grammar often, and a very small percentage of students said they never learned grammar in Croatian language lessons.

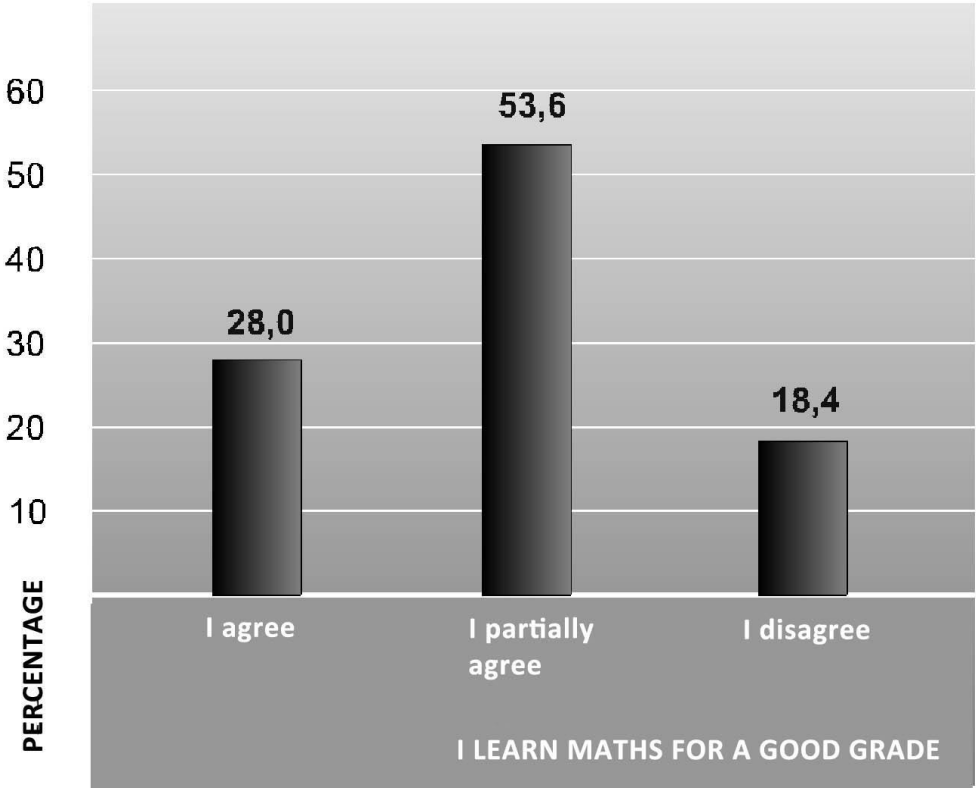
7.3.3. Maths

The part of the questionnaire for Maths (PU-2) consists of 24 questions and it refers to curricular and extracurricular experiences of the students connected to Maths. This part of the *Report* deals with the results of the background questionnaire connected to Maths.

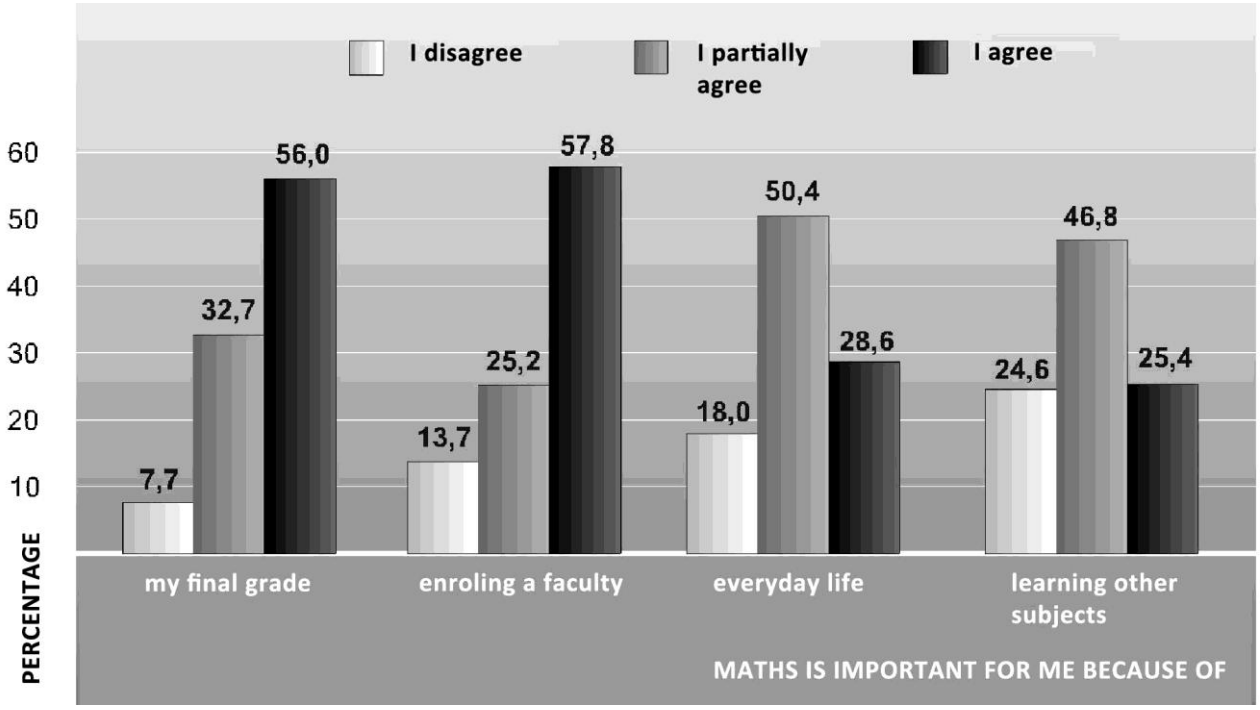
The response interpretation refers to the students attitudes toward some aspects of *motivation to learn Maths*, experiences in *learning contents of Maths in curricular and extracurricular activities*.

Motivation. A number of questions was aimed at yielding information on students motivation to learn Maths, i.e. to what extent they learned Maths for a good grade. Graph 18 shows that more than 50% of students thought that the grade was not the only reason they learned Maths. More than one third of students considered the grade the only reason to learn Maths, whereas less than one fifth of the students did not agree with the statement that they learned Maths because of the grade. Responses in both subject show quite similar responses in the field of motivation.

Graph 18. Motivation to learn Maths



Graph 19. Motivation to learn Maths



Graph 19 further explores other types of motivation to learn Maths. More than 50% of students thought learning Maths was important to enrol a university and to get a good grade at the end of the school year. About one fourth of the students thought learning Maths was important for their everyday life and for learning other subjects, and about a half of the students partially agreed with the statements. The last fourth did not agree with the statements that learning Maths was important for learning other subjects.

Motivation was further explored with a question how interesting mathematics as a subject was. Less than half the students (42,2%) thought mathematics was partially an interesting subject. More than a fifth of the students (20,9%) thought agreed with the statement that mathematics was an interesting subject, whereas one fifth of the students (33,4%) did not consider Maths interesting at all.

Maths classes. Questions connected to Maths lessons refer mainly to the teaching methods used and individual lesson contents and how they relate to everyday life, to the use of teaching materials and aids and to some of the phases of the teaching process.

One of types of learning mathematics is learning by exploring: Majority of students (60.2%) never learned Maths by exploring, 29.4% of them sometimes, and 4.3% of the sample students reported that they often learned Maths by exploring.

Table 25. Experiences in Maths classes

	Never	*Sometimes*	*Often*
We work in groups in Maths classes.	71.2 %	21.9 %	3 %
I learn Maths contents by exploring.	61.2 %	29.4 %	4.3 %
We repeat lesson contents of writing in Maths lessons.	5.8 %	31.3%	35.7 %
There is enough time to practice exercise in the Maths classes.	3.6 %	18.5 %	32.6 %

Repeating the lesson contents and doing exercises are important phases of the teaching process. *Table 26* depicts that majority of students often and always repeats the previously learned contents in Maths classes, whereas one third of the students reported this to occur sometimes. For students to learn a certain lesson contents, they need practice. Therefore, they are asked whether they had enough time during classes to do practice exercises. Most of the students said they had enough time during classes to do practice exercises, almost one fifth said they sometimes had enough time during classes to do practice exercises, whereas a small percentage of students said they did not have enough time during classes to do practice exercises.

A special attention in exploring experiences connected to Maths classes was paid to **students' activities during classes and types of work**. These students' experiences were also explored with the scale consisting of four elements: *always, often, sometimes, and never.*

13,1% students said they actively took part in Maths lessons, often was reported by 29,8% of the students, sometimes by 44,7% of them, and by 9,0% of them. 23,4% of the students completely agrees with the statement that their teacher has an interesting was of teaching Maths, and 35% did consider Mathematic lessons interesting. 38,5% o the students consider Maths only partially interesting.

19,6% of the students completely agrees with the statement that their teacher teaches Maths comprehensibly, and 29,7% of the students said that their Maths teacher often taught Maths comprehensibly. 33,2% the students completely agrees with the statement that their teacher teaches Maths in an interesting way, 33,2% of the students said this occurred sometimes, whereas 13,7% said their Maths teacher never thought in an interesting way.

36% of students reported they often got **feedback on their performance during the classes**, 43,6% students reported this to occur sometimes, whereas 16,9 % of students said they never got feedback on their performance during the classes.

In Maths classes 59,6% of the students said they sometimes used **example from everyday life**, 23,7% reported never to use them, and 13,5% of the students said they often or always used these examples.

The teaching content of other subjects always or often correlated with the teaching content of Maths in 30,3% cases, and 55,8% of the students sometimes had this possibility. 9,6% of the students said they never correlated teaching contents of Maths with the teaching contents of other subjects.

During Maths classes **Maths lesson contents often or always correlate with the content of other subjects** in 9% of the reported cases, sometimes in 58.3%, and never in 29,4% of the reported cases.

Teachers are encouraged to use different teaching aids to help their students better comprehend the Maths contents. *Table 26* shows the responses to the questions about used teaching aids.

Table 26. The use of teaching aids in Maths lessons

	We do not use it so I cannot say	I disagree	I partially agree	I agree
overhead projector	68,2 %	16,2 %	10,3 %	2 %
geometric models	44,6 %	6,5 %	29,4 %	16,2 %
computer	55,3 %	8,3 %	17,3 %	15,9 %
smart board	59,4 %	4,5 %	12,8 %	8,5 %

The most frequently used teaching aid are geometric models which almost half the students use. 16,2% of them completely agrees with the statement that the use of geometric models helps them better understand Maths content. 29,4% of the students partially agrees with the statement, and a small percentage did not agree with the statement.

Majority of students said they never used overhead projectors in their classes, and those who did (16,2%) mostly did not agree with the statement that it helped them better understand Maths content. Only a small number of students completely or partially agreed with the statement.

44,7% of the students use computer during the lessons, and 40,6% use the smart board. 5,9% of the students find the computer a useful tool. 8,5% of the students thought that about the smart board. Only a small number of students found neither computer nor the smart board a useful aid in helping them better understand Maths content.

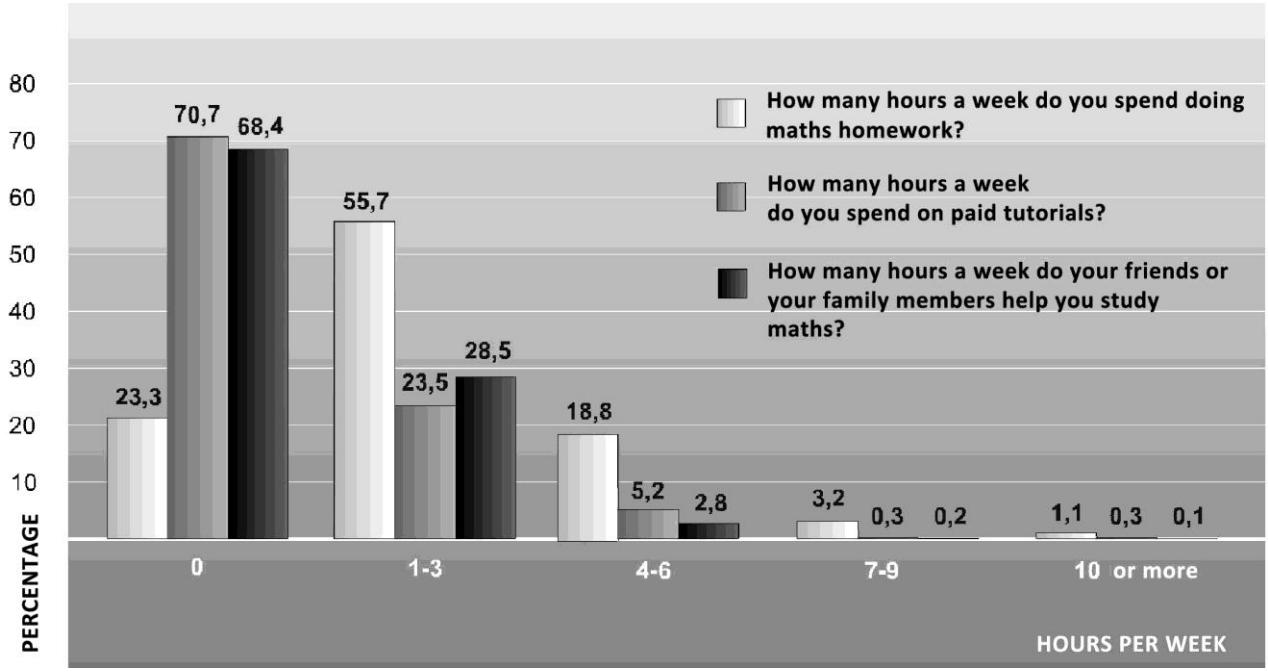
Learning Maths in extracurricular activities. Questions related to this area referred to doing homework, taking lessons and other forms of learning aids.

The most frequent form of the extracurricular activity is doing homework. This is also confirmed by the responses students gave to the question about the frequency of getting homework. 60,6% of students have homework to do after every Maths lesson, 25.5% of them said they sometimes got homework, whereas 2.1% said they never got homework. Despite the fact that a great majority of students (86,5%) reported they were regularly assigned exercises to do at home, 20.5% never did that.

Graph 20 shows responses connected to these areas. It can be seen that more than half of the students spend on average between one and three hours a week doing Maths exercises. One fourth of the students spends between four and six hours a week doing Maths exercises, 3% of the students spends between seven and nine hours a week doing Maths exercises, and only less than 1& spends more than 10 hours a week doing homework.

Students sometimes require help in doing their homework. About 27% of students take paid private classes, and 33% of them have their family members or their friends help them.

Graph 20. Learning Maths in extracurricular activities



8. PSYCHOMETRIC ANALYSIS OF SCORES OF NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENTS

Natalija Ćurković, Josip Šabić

The purpose of the psychometric analysis was to establish the level of reliability and validity of each test booklet as well as the each item's characteristics. This analysis is necessary to ensure that the obtained test results indeed represent the level of knowledge of certain teaching contents. Therefore, the group in charge of psychometrics ran the psychometric analysis of all test booklets after pre-testing. The feedback on each item's characteristics as well as on the whole test booklets was very useful to the Croatian language group and for the Maths group in their effort to improve their test booklets for the main test administration. In other words, based on the information on psychometric characteristics of items and test booklets the work groups were able to decide which items they were going to use for the main test administration.

After the test in Croatian language and Maths had been administered, the psychometric analysis was rerun based on *classical test theory* (CTT) and *item response theory* (IRT). *Classical test theory* is the body of related psychometric theory that predicts outcomes of testing, such as the difficulty of items, their discrimination and reliability (Allen and Yen, 2002). *Item response theory* is a modern system of mathematical models that describe the connection between the latent trait and its manifestations (Ayala, 2009). The student's knowledge is in this analysis the latent trait whereas the student's achievement on the test aimed at testing that knowledge is its manifestation. These two approaches have different starting points and methods, but the end result is the same - determining whether the items and the test fulfil the requirements. Combining the two approaches, the working group for psychometrics intended to gain detailed insights into the test booklets' and items' characteristics.

8.1. PSYCHOMETRIC ANALYSIS ACCORDING TO CLASSICAL TEST THEORY

The mean (M) is the sum of all scores, divided by the number of students. If the test is appropriate for a certain group of students, the mean should be placed at the half of the possible range. If the mean is pulled up or down, the test is either too easy or too difficult for that group of students, which means that that test cannot be used to fully distinguish between the students.

The range, as another indicator of a test's appropriateness, is the difference between the top and the bottom score achieved on the test, and it shows the scope of obtained numerical values in the set of the scores, and it can also be used as an approximate orientation scatter score index (Field, 2005). A test is expected to be designed for the students to achieve the maximum range (from zero to the maximum possible score) since only this range enables the test to best distinguish between higher and lower level students. In case none of the students achieves the maximum score, it would mean that the test was not appropriate for the targeted group of students. In other words, it was too difficult. However, other explanation for this could be that the students taking the test were not motivated enough to achieve their best at the test.

The problem is if the bottom score is far away from zero. This would mean that the test had too many easy items that all students answered correctly and it did not distinguish between higher and lower level students.

Standard deviation is the average amount that each student's score deviates from the mean. It shows the density of scores' distribution around the arithmetic mean. (Petz, 2005).

Standard error of measurement is an estimate of error to use in interpreting an individual’s test score, and it can

be calculated from the level of reliability. It is used to estimate the true score by constructing a range within which the student's true score is likely to fall given the obtained score. (Petz, 2005). This is a very important factor that raises our awareness of the possible error we might make when we make conclusions about the test scores. This must be taken into account when ranking students by their scores. If a test has a high standard error of measurement, simple ranking students by their score might be incorrect and it might disadvantage a student.

Reliability is the internal consistency of a test measurement. It measures the average correlation between all test items and it is expressed by Cronbach's alpha coefficient. Cronbach's α coefficient is a correlation coefficient and as such it varies between 0 and 1. Different levels of this coefficient are acceptable for different types of measurements. For this kind of testing it should be at least 0,90 (Kehoe, 1997). However, Cronbach's α coefficient depends on the number of items - the greater the number of items, the more reliable the test is, i.e. Cronbach's α coefficient is higher. Tests with a low number of items it cannot be expected to have high values (≥0,90).

Item difficulty is a measure of the proportion of examinees who responded to a dichotomous item correctly. This measure does not specify whether an item is good or bad, but it merely indicates how difficult the items for a certain group of students was (Osterlind, 2001). In case of not dichotomous items, the item difficulty can be calculated by dividing the arithmetic mean by the maximum score for that item.

For a test to be of appropriate difficulty, at least half the items should be of medium difficulty, i.e. items that can be done correctly by 40% to 60% of students. The other half should be consisted of difficult and easy items evenly distributed, i.e. there should be an equal number of difficult and easy items. However, there should not be too many difficult or too many easy items. The items done correctly by 10% of students are considered to be difficult, whereas the items done correctly by 90% of students or more are considered to be easy. A certain percentage of such items is necessary to be able to discriminate between very high and very low level students, but the whole test should not contain more than 10% of items of both categories (Državni ispitni center, 2007). In doing so it is crucial for the numbers of easy and difficult items should be the same.

Discrimination or discriminative value is a measure of how an item discriminates between students who are knowledgeable in the same content area and those who are not (Haladyna, 2004). Students who do a discriminative item correctly will most probably do well on the whole test. Thus it can be said that this feature directly shows an item's quality (Osterlind, 2001). Discrimination or discriminative value is expressed by discrimination index (DI) calculated as the correlation of an item's correctness and the overall test's score if the item is excluded from the overall test's score (Norusis, 1998). The calculated correlation should be as high as possible. High DI means the higher correlation between the item and the test's score, but also that the particular item distinguishes well between the students in view of their knowledge. Low DI (around zero) means that this correlation is accidental and these items should be avoided. Negative correlation means that low level students do an item right whereas high level students experience difficulty when doing the item. Such items must be moderated.

The lowest acceptable DI is 0,2 (Tucker, 2007), and a well-designed test should not have more than 20% items with DI lower than this minimum value. (Državni izpitni center, 2007).

Table 27 contains parameters obtained according to classical test theory, and they refer to the analysis of each test booklet. The analysis was run by TiaPlus software developed by Dutch experts from CITO-a and specially adjusted for running analysis according to *classical test theory* and for test specifications with more test booklets (Heuvelmans, 2008).

8.2. PSYCHOMETRIC ANALYSIS ACCORDING TO ONE-PARAMETER MODEL *OF ITEM RESPONSE THEORY*

Basically, any approach of the item response theory supposes that in a testing situation student's performance on the test can be predicted (or explained) by defining attribute found in the basis of the performance. The relation between the "observable" and the "latent" quantities is described by a mathematical function. For this reason, IRT models are mathematical models based on assumptions about the test data. Mathematical function describing an item is a characteristic curve of the item defined by three parameters: item difficulty, discrimination index, i.e. curve steepness and pseudo-guessing. Depending on the number of parameters used to describe the function, the models are usually referred to as one-, two- or three-parameter models.

When analysing the scores a one-parameter model was used. This model is very often called the Rasch model after the Danish Mathematician G. Rasch who developed it. It uses only item difficulty when defining the item characteristic curve. However, the model used in this analysis actually is a modification of the Rasch model done by the Dutch experts. This modification refers to the possibility of a retrospective addition of discrimination parameters. Based on this approach a new software OPLM One-Parameter Logistic Model was used in this analysis (Verhelst, Glas i Verstralen, 1995). The use of the one-parameter logistic model is recommended with a smaller sample (Hambleton, 1994). Test in Croatian language was taken by 990 students, and in Maths 1244 students. However, due to the fact that a large number of students taking the test in Maths who did not provide any response, all students who left 25% or more of items undone were not included in the analysis. About 25% of students were not included in the analysis. This number would normally be considered big enough, but due to the application of the incomplete test specifications in the research, not all of the students did all the items. Eventually, the final range of the students whose scores were analysed was between 140 and 530 in Maths, whereas about 660 students did almost all items in Croatian language. In terms of item response theory samples of 500 to 600 students per item belong to the group of minimum samples to run the analysis (Hwang, 2002).

Table 27. Psychometric features of scales of levels of knowledge subject domains of Croatian language and Maths obtained *by classical item theory*

TEST BOOKLET	N	Number of items	O	C	M	SD	SEM	M(ID)	ID <1	ID >0.9	ID %	RI(α)	α (40)	M(DI)	DI%	Range	Max
Croatian language																	
IKHJ-1	329	51	24	27	32.3	10.10	4.16	0,49	2	-	3,9	0,83	0,79	0,33	29,4	0 – 66	66
IKHJ-2	333	69	24	45	47.9	11.62	4.51	0,53	3	2	7,2	0,85	0,77	0,30	39,1	0 – 90	90
IKHJ-3	328	64	22	42	45.9	12.24	4.53	0,53	3	2	7,8	0,86	0,80	0,32	39,1	0 – 86	86
Maths																	
IKM-1	180	48	39	9	23,5	11,15	2,89	0,46	1	-	2,1	0,93	0,92	0,50	-	0 – 51	51
IKM-2	170	48	47	4	22,4	10,66	2,91	0,44	3	-	6,3	0,93	0,91	0,48	4,2	0 – 51	51
IKM-3	178	48	44	4	22,0	9,84	2,82	0,46	5	-	10,4	0,92	0,90	0,46	2,1	0 – 48	48
IKM-4	178	48	40	8	22,8	9,49	2,87	0,45	3	1	7,8	0,91	0,89	0,44	6,3	0 – 51	51
IKM-5	181	48	44	4	22,0	10,04	2,88	0,43	2	-	4,2	0,92	0,90	0,47	-	0 – 51	51
IKM-6	179	48	39	9	20,6	8,66	2,86	0,43	2	-	4,2	0,89	0,87	0,41	4,2	0 – 48	48
IKM-7	178	48	37	11	22,2	9,57	2,84	0,46	2	-	4,2	0,91	0,90	0,44	4,1	0 – 48	48

Legend

N - number of students doing the test booklet
Number of items - number of items in the test booklet

O - open-ended questions

C - closed questions
M - arithmetic mean of the performance on the test booklet
SD - standard deviation of the performance on the test booklet
SEM - Standard error of measurement of the performance on the test booklet
M(ID) - average index of difficulty

ID<0.1 - number of items with the index of difficulty less than 0.1

ID % - number of items with the index of difficulty less than 0.1and greater than 0.9 (there should not be more than 10%)
RI(α) - reliability index of the test booklet (Cronbach α coefficient)
α (40) - estimate of Cronbach α coefficient in case the test booklet contains more than 40 items
M(DI) - average discrimination index of the test booklet
DI% - percentage of items with the discrimination index (DI) less than 0.2%

(there should not be more than 10%)

Range - the range of scores students achieved on the test
Max - maximum score the students could obtain on the test - the analysis was not run due to

ID>0.9 - number of items with the index of difficulty greater than 0.9 small number of students

Table 28. Psychometric features of scales of levels of knowledge subject domains of Croatian language and Maths obtained *by item response theory*

Subject domain	Number of items		DI		Distribution of p-values for S_i-tests												R_{1c} test			Number of students	
	before	after	range	GM	≤ ,05	,1	,2	,3	,4	,5	,6	,7	,8	,9	1,0	?	R_{1c}	ss	p	M	Range
CROATIAN LANGUAGE																					
G	22 (15)	19 (12)	1 – 4	1,7	2	1	2	3	4	1	0	1	0	3	1	1	135,40	126	,27	661	
RLT	17 (15)	17 (15)	2 – 5	2,6	1	1	2	1	2	1	3	2	2	0	2	-	128,41	108	,09	660	657 – 662
RNT	28 (20)	26 (19)	1 – 4	1,8	1	0	5	3	4	5	2	0	1	3	2	-	197,69	171	,08	660	657 – 662
TL	26 (15)	26 (15)	1 – 6	2,5	2	2	2	0	4	1	3	5	1	3	3	-	173,82	170	,41	662	661 – 662
MATHS																					
F	24	24	1 – 5	2,8	2	1	4	1	1	2	0	3	1	5	4	-	124,33	117	,31	269	141 - 281
E	27	26	1 – 3	2,0	2	5	0	2	4	3	6	0	2	0	2	-	152,00	138	,20	253	248 – 260
L	22	22	1 – 4	2,0	0	0	3	2	4	0	5	2	2	2	1	1	150,47	128	,08	275	269 – 281
C	27 (24)	27 (24)	1 – 4	2,6	5	1	7	3	3	1	2	1	2	2	0	-	311,76	277	,07	465	398 – 532
F	24	24	1 – 3	1,9	0	2	2	1	5	3	1	3	1	2	4	-	156,99	142	,18	275	269 – 281
T	27	27	1 – 2	1,1	4	0	3	1	2	2	6	2	4	1	2	-	138,99	127	,22	253	246 – 260

Legend
Number of items - before - the total number of items in the subject domain (in front of the brackets is the number of tasks if each point in polytomous tasks is treated as a

p - the level of statistical significance
M - average number of students per task
G - grammar

separate item; in the brackets is the number of tasks if each point in polytomous tasks is not treated as a separate item
Number of items - before - the total number of items in the subject domain after calibration
SD - standard deviation of the performance on the test booklet
ID - discrimination index
GM - geometrical mean of discrimination indices
? – p value outside the range
R1c - index of fitting of the item to the model
ss - number of degrees of freedom

RLT - reading literary texts
RNT - reading non-literary texts
TL - theory of literature
F - quadratic function
J - quadratic equation
L - linear equation
C - Mathematical literacy
F - fractions
T- triangle

Table 28 contains results obtained according to one-parameter model of *item response* theory, considering the subject domains.

Before running the analysis according to *item response theory* it is necessary to find a mathematical model that best describes (fits) the items. Looking for the most appropriate model is done by many calibrations which include a very comprehensive work that cannot be described here. Only when such a model is found, the obtained parameters are explained. During the calibrations, i.e. the search for the most fitting models, it is sometimes necessary to omit an item that does not fit the model used to explain the functioning of a particular subject domain. The intention is to find a model that will require omission of the smallest number of items. Hence, there are two columns with the item numbers in *Table 28*. The first number is the one the subject domain contained, and the other one is the number after the omission of some items in the process of calibration.

Table 28 further contains columns with geometric means and the range of discrimination index (DI) of particular subject domain. Discrimination indices describe how well each item distinguishes students based on their abilities to do that particular item (Baker, 2001), i.e. how well the item discriminates between students with higher and those with lower ability levels (Ayala, 2009). The ability level is defined by the criterion of over 50% probability of getting the item right (Ljubotina, 2000). Students who have less than 50% chance of getting an item right belong to the group of students with low ability level, and vice versa. The range of discrimination indices refers to the discrimination indices in particular subject domains. The higher the value, the higher the discrimination index. The average item discrimination in particular subject domain is derived from the geometric mean which is the average ratio of discrimination indices (Petz, 2005).

As already mentioned, the purpose of calibration is to find the model that best fits the obtained scores. Si-tests are conducted to establish the deviation of some items from the suggested model. The end results of calibration as well as of the Si-tests for the accepted models in *Table 28*. The results of Si-tests are expressed in p-values which are expected to be evenly dispersed within the different intervals from 0 to 1. Lower p-values mean greater deviations from the model, whereas greater p-values increase the probability o a model being accepted for the subject domain (Verhelst, Glas i Verstralen, 1995).

R1c is a rough test considered to be a combination of Si-tests (Verhelst, 1993). It can also be compared to s2 χ(hi-square) test, i.e. it evaluates to what extent the suggested model differs from the obtained results. If the model statistically differs from the data, R1c gets p-values less than 0.05. The aim of the calibration is to find a model that statistically speaking does not significantly differ from the results, i.e. to obtain p-value greater than 0.05. The obtained values can be seen in *Table 28.*

REFERENCES

Allen, M. J. i Yen, W. M. (2002). Introduction to Measurement Theory. Long Grove, IL: Waveland Press.

Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.

Baker, F. B. (2001). *The Basics of Item Response Theory*. University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Braš Roth. M. (ur.) (2008). Pisa 2006 *Prirodoslovne kompetencije za život*. Nacionalni centar za vanjsko vrednovanje obrazovanja – Pisa centar, Zagreb.

Cohen, L. i drugi (2007). *Metode istraživanja u obrazovanju*. Jastrebarsko: Naklada Slap.
Državni izpitni center. (2007). *Letno poročilo. Splošna matura 2007*. Ljubljana: Državni izpitni center.

Europska komisija,Uprava za obrazovanje i kulturu (2004). Provedba radnog programa *Obrazovanje i izobrazba i izobrazba 2010.,* Radna grupa B, *Ključne kompetencije, Ključne kompetencije za cjeloživotno učenje, Europski referentni okvir, Europska komisija, Uprava za obrazovanje i kulturu,izabrala* Nevenka Lončarić Jelačić, prevela Zlata Pavić, preuzeto s mrežnih stranica Agencije za odgoj i obrazovanje www.azoo.hr *, na dan 15. srpnja 2009.*

Field, A. (2005). *Discovering Statistics Using SPSS.* London: Sage.

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K. (1994). Item Response Theory: A Broad Psychometric Framework for Measurment Advances. *Psicothema, 6(3),* 535-556.

Heuvelmans, T. (2008). *Tia Plus Users Manual.* Measurment and Research Department. Arnhem: CITO.

Hwang, D. (2002). Classical Test Theory and Item Response Theory: Analytical and Empirical Comparisons. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX.

Kehoe, J. (1997). *Basic Item Analysis for Multiple-Choice Tests.* http://www.ericdigestd.org/1997-1/basic.html

Ljubotina, D. (2000). *Usporedba psihometrijskih karakteristika kompozitnih testova konstruiranih u sklopu klasične teorije i teorije odgovora na zadatak*. Neobjavljena doktorska disertacija. Zagreb: Odsjek za psihologiju Filozofskog fakulteta u Zagrebu.

Ministarstvo kulture i prosvjete Republike Hrvatske (1994). *Nastavni programi za gimnazije*, Zagreb: Ministarstvo kulture i prosvjete Republike Hrvatske. 130

Ministarstvo znanosti obrazovanja i športa, Zagreb (2005). *Plan razvoja sustava odgoja i obrazovanja* 2005. – 2010. Zagreb: Ministarstvo znanosti obrazovanja i športa.

Norusis, M. J. (1998). *SPSS/PC+ Advanced Statistics V2.0 for the IBM PC/XT/AT and PS/2.* Chicago: SPSS Inc.

Osterlind, S. J. (2001). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats.* Boston: Kluwer Academic Publishers.

Petz, B. (2005). *Psihologijski rječnik.* Jastrebarsko: Naklada Slap.
Povjerenstvo za standarde pedagoškog i psihološkog testiranja AEA, APA i NCME (2005). *Standardi za pedagoško i psihološko testiranje.* Jastrebarsko: Naklada Slap.

Primorac, D. i drugi (ur.) (2008). *Ususret državnoj maturi*, Zagreb: Ministarstvo obrazovanja i športa Republike Hrvatske, Zagreb.

Tucker, S. (2007). *Using Remark Statistics for Test Reliability and Item Analysis.* Neobjavljeni rad. Baltimore: University of Maryland.

Verhelst, N. D. (1993). *On the standard errors of parameter estimates in the Rasch model.*

Measurment and Research Department Reports, 93-1. Arnhem: CITO.

Verhelst, N. D., Glas, C. A. W. i Verstralen, H. H. F. M. (1995). *OPLM: One Parameter Logistic Model. Computer program and manual.* Arnhem: CITO.

SAŽETAK

Izvješće o projektu *Razvoj i strategija nacionalnih ispita* iznosi ciljeve, aktivnosti i rezultate *Projekta*. Cilj je *Projekta* bio educirati djelatnike *Nacionalnog centra za vanjsko vrednovanje obrazovanja* i njegove vanjske suradnike u području vanjskoga vrjednovanja. U okviru edukacije provedeni su *nacionalni ispiti– nacionalna procjena postignuća* na uzorku učenika drugoga razreda škola s gimnazijskim programom.

U *Izvješću* su opisani rezultati izradbe testova iz hrvatskoga jezika i matematike, počevši od odabira sadržajnih područja ispitivanja i obrazovnih ishoda pa do same izradbe ispitnih zadataka i oblikovanja nacrta testova. Isto je tako opisan proces izradbe popratnih upitnika od osmišljavanja ciljeva preko sastavljanja pitanja do opisa rezultata. Rezultati *nacionalnih ispita – nacionalne procjene postignuća* analizirani su prema *klasičnoj teoriji testova* i *teoriji odgovora na zadatak*. Ovim su *Projektom* prvi put provedeni formativni ispiti u Republici Hrvatskoj čime su postavljeni temelji za razvoj nacionalnih ispita s ciljem poboljšanja kvalitete hrvatskoga obrazovnog sustava.

SUMMARY

This report describes aims, activities and results of the project *Development of Instruments in Croatian National Assessment*. The aim of this *Project* is capacity building in educational assessment knowledge of *National Centre for External Evaluation of Education* staff and external experts. In the scope of this *Project,* National Assessment of student achievements was conducted on the 2nd grade gymnasium program students' sample.

Results of the test development for Croatian Language and Maths is described in this report, which includes subject domains, educational outcomes, item writing and test design development. The process of Background Questionnaires construction is described regarding to aims, structure and description of results. Results of National Assessment were analyzed according to Item Response Theory and Classical Test Theory. Through this Project first formative assessment of students achievement was conducted in Croatia and foundations of an external assessment system were set to ensure quality in education.