ISSN 2278-3091
Volume 8, No.1.1, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse4981.12019.pdf
https://doi.org/10.30534/ijatcse/2019/4981.12019

Analysis and Comparison of the National Diagnostic Paper-Based and Online Tests

Loreta Juskaite
Riga Technical University, Kalku Street 5,
Riga, Latvia, LV-1658,
Loreta.Juskaite@rtu.lv

ABSTRACT

A National diagnostic test is one of the diagnostics tools used to ascertain students' basic skills acquisition on finishing a certain level of education according to the national education standards and the subject curriculum requirements. The rapid development of information technology (IT) contributes to conducting and assessing national diagnostic tests online at the same time increasing its role in the modern education system. The aim to examine whether the traditional paper-based tests are applicable in a different environment (in an e-environment Online/Computer-based test), whilst maintaining its theoretical and technical specific features and is to focus on the sustainable development of activities in order to improve the quality of education in STEM (Science, Technology, Engineering and Mathematics) subjects, which includes using innovative methods in the Latvian educational context.

Key words: Information technologies (IT), Online /Computer-based test, National diagnostic (ND) test, Cognitive activity level

1. INTRODUCTION

The rapid development of IT contributes to the increasing use of e-environment in the modern education system. [1] Is this process worth implementation in an era when computer and e-environment became a familiar source of information for many young people and should National diagnostic tests be conducted only in an e-environment? Are there possible combined variants? Should ND tests proceed the traditional way or are there major changes needed? The aim of national diagnostic and standardized tests is to find out students' basic skills acquisition on finishing a certain level of education according to the national education standards and the subject curriculum requirements, as well as to enable local authorities, schools, teachers and students to independently and objectively evaluate learning achievements, obtaining detailed feedback in order to evaluate the results and to improve the learning process. [2, 3, 4]

In 2015/2016 and in 2016/2017 Latvian Secondary schools were offered the opportunity to choose how to organize the diagnostic test – paper-based or Online/computer-based. [5, 6] In order to obtain valid and reliable data it was important to involve educational institutions if possible from all regions of the country, as well as all types of educational institutions. Paper-based diagnostic tests in physics were conducted at the same day and time. This made it possible to research whether traditional paper-based diagnostic tests can be conducted in a digital form online, thus promoting the use of innovative IT to examine students' knowledge. [7, 8, and 9] At the same time online computer-based tests modernize and facilitate the assessment and analysis process, as well as ensure sustainable elaboration of diagnostic tests and their customization to the new STEM model requirements. [3] The analysis of the results enables the search for new approaches and methods, for new information systems and technologies in order to enhance certain skills, abilities and attitudes.

After conducting the diagnostic tests tasks teachers took part in a survey about the computer-based diagnostic tests. More than 98% of the surveyed teachers evaluated the online diagnostic test very positively. Teachers and school administrations gave a positive evaluation of such online tests' ad-vantages as:

1) *Automatic test assessment and data transfer to the National Centre for Education*

2) *Announcement of the results immediately after fulfilling the test e*

3) *Saving of time and school resources [6]*

Problem – if a diagnostic tool created in a traditional way for the paper-based environment is applicable in a different environment (in Online/ Computer-based), whilst maintaining its theoretical and technical specific features and ensuring consistent and comparable measurement of results. The research aims – to justify the adaptation of diagnostic tools created in a traditional environment in the Latvian educational context.

Structure of this article: scientific literature analysis summary, quantitative research using ITEMAN Test Analysis Program, qualitative and quantitative analysis of the research findings.

2. THEORETICAL BACKGROUND

ND tests make it possible to obtain valid and reliable data on acquisition of subject standard requirements, which facilitates estimating students' readiness to master a particular subject. This is one of the essential tools that contribute to the implementation of national objectives set out in the development plan documents. It also helps teachers to under-stand subject standards, to plan teaching activities and to evaluate its outcomes. In order to achieve high results it is important to invest in purposeful work. [10]

ND tests also help teachers understand the meaning of the standard requirements acquisition on a higher cognitive level (according to the OECD PISA the level 5 and level 6). ND

tests contain tasks to examine knowledge and skills, the acquisition of which is required for the successful continuation of education. [2, 6, 9, 10, and 11]
The conceptual document "Latvian Growth Model: People First" (approved by the Latvian Government in October 2005) contains one of the priority long-term tasks - an increase in the proportion of STEM subjects learning at all levels of the education system.

2.1 The situation in the Latvian Education System. Characteristics

A global education survey of the OECD PISA1 2006, 2009, 2012 and 2107 results show that in Latvia there are a relatively small proportion of students with high achievements in reading, mathematics and science, and that number continues to decline. In the information report of the Minis-try of Education and Science "The planned courses of action and measures for improvement of learning and teaching quality in mathematics, sciences and engineering and for in-creasing the amount of trained specialists" it is stated that the most important aspect is the improvement of students' reading literacy, mathematics and science competencies, ensuring the achievement of the goals declared in the Latvian National Development Plan 2014 – 020 (lowest level of reading literacy in 2017 - 15%, in 2020 - 13%). [9, 10, 11]

2.2 Education Development Guidelines in the Latvian Education System

Education Development Guidelines for 2014 - 2020 specify the percentage increase in reading literacy, mathematics and science from 4.2%, 8% and 4% in 2012 to 7 %, 8% and 8% in 2020 among the students with high educational achievements (OECD PISA Level 5 and Level 6). According to the OECD PISA scale in science for 2015, students who have reached the competence level 6 in science are able to use content, procedural and epistemic knowledge to consistently provide explanations, evaluate and design scientific enquiries and interpret data in a variety of complex life situations that require a high level of cognitive demand. They can draw appropriate inferences from a range of different complex data sources, in a variety of contexts and provide explanations of multi-step causal relationships. [11, 12]

2.3 Classical Test Theory in National diagnostic tests

National diagnostic tests are different from other tests, because they are developed using special methods by selecting the tasks that should meet certain criteria and are designed for measuring students' learning achievements. Education Development Guidelines also highlight a need for the monitoring of learning achievements at the national, school and class levels to obtain objective data about students' knowledge and skills, and to establish a causal link between the factors influencing learning process and students' achievements, and to improve planned measures for better learning achievements in the educational content and process. [13, 14]

For data analysis a classical measurement procedure was used, which is based on the Classical Test Theory (CTT). Classical Test (CT) analysis is a tool to measuring individual differences. CTT introduces three basic measurement concepts: *Test score or observed score, True score and, Error score.*

CT analysis postulates linking the observed test score (X) to the sum of the true score (latent unobservable score) and error score: $X = T + E$. The following assumptions underlie CTT:

1) *True scores and error scores are uncorrelated.*
2) *The average error score in the population of examinees is zero.*
3) *Error scores on parallel tests are uncorrelated.*

CT analysis utilizes traditional item and sample dependent statistics. Classical test analysis also typically includes a measure for the reliability of scores (i.e., Cronbach Alpha) and difficulty of the test. [15-21] Several benefits are obtainable through the application of good instructional objectives and item writing using CT analysis. [19]

When compared to item response theory models, analyses can be performed with smaller representative samples of examinees. This is particularly important when field-testing a measuring instrument. CT test analysis employs relative simple mathematical procedures and model parameter estimations are conceptually straightforward. CT analysis is often referred to as “weak models” because the assumptions are easily met by traditional testing procedures. [21]

2.4 Test Analysis Program

Results were obtained analysing the diagnostic tests with ITEMAN Test Analysis Program ITEMAN™ for Windows. The ITEMAN software program analyses multiple-choice questions and can also compute and record test scores. The offers four statistical measures: 1) Ease Index (called ― Proportion Correct“), 2) Discrimination Index, 3) Biserial.4) Point Biserial Correlation Coefficients.

ITEMAN requires that the input data file be formatted in ASCII (text-only) files. The output file contains the statistical measures, and displays them not only for each question, but for each alternative as well. ITEMAN is easier and faster to use. [23, 24]

3. NATIONAL DIAGNOSTIC TESTS RESULTS-NEEDS AND DESIRES

In accordance with the requirements of the Cabinet of Ministers Regulations Nr.468 "Regulations on the national education curriculum, education subject curriculum and education subject programs" of the 12 August 2014 and physics subject curriculum "Physics Grades" (Annex. 10) the diagnostic test aims to obtain detailed feedback, to assess students' knowledge and skills in order to improve the teaching and learning process of physics at a secondary education level. The technical support of the online diagnostic tests provided the portal uzdevumi.lv, thereby contributing to the IT use examining students’ knowledge and facilitating the assessment of tasks and analysis process of the results. [6, 9]

3.1 Methodology for diagnostic tests' aspects

Overall, the diagnostic tests' results were compared in various aspects: On the basis of school teaching language, On the basis of the type of school, on the level of urbanization.

According to the collected data, different types of educational institutions (secondary schools, gymnasiums, state gymnasiums, technical schools, vocational schools) from virtually all regions of the country were represented as shown in table 1. In the analysis diagnostic tests performances of 10 948 students were used.

3.2 National diagnostic tests results

Table 1: The Number of Schools and Students Involved (2015/2016)

The number of schools		The number Of students	Percentage
Computer-based online tests	30	2 403	46%
Paper-based tests	62	2743	54%
Total	92	5146	

In 2016/2017 3059 students' performances in e-environment (Computer-based) which is 656 or 13% more than in the previous school year [6], as shown in table 2

Table 2: The Number of Schools and Students Involved (2016/2017)

The number of schools		The number of students	Percentage
Computer-based online tests	33	3059	53%
Paper-based tests	59	2743	47%
Total	92	5802	

3.3 Description of the diagnostic test

The diagnostic test is composed of one variant and consists of 35 multiple-choice tasks. Students have to choose one correct answer from four options. Task content is selected according to a certain level of the mandatory learning content in physics. In order to assess students' achievements, the test contains tasks with different difficulty and cognitive levels, content of which relatively covers 7 aspects contained in the education curriculum in physics [6, 11] as shown in table 3.

I

Table 3: The division of tasks according to the skills and cognitive Level

Nr.	Curriculum aspects (topics)	*Cognitive level*			*Total*	
		Memorization and understanding (knowledge)	*Usage of knowledge and skills*	*Analysis and productive activity*	*The number of tasks*	*Task percentage*
1	Physical terms and units of measurement	4	1	0	5	14,3%
2	Understanding and usage age of physics concepts	3	2	0	5	14,3%
3	Interconnect-tions of the physical quantities	0	4	1	5	14,3%
4	Explaining an experi-ment	0	5	0	5	14,3%
5	Understand-ing processes	0	5	0	5	14,3%
6	Graphical representa-tion of phys-ical process-es, circuit diagrams, vectors		4	1	5	14,3%
7	Scientific enquiry	1	2	2	5	14,3%
To tal	The number of tasks	8	23	4	35	
	Task per-centage	23%	66%	11%		100%

First-level cognitive activity tasks correspond to one step of the operation or procedure. For example, the student remembers the concepts, definitions, physical quantities, physical units, interconnections or principles, recognizes characteristics, devices or processes.

Second-level cognitive activity tasks focus on physical phenomena and understanding of processes usage. Whereas third-level cognitive activity tasks make it possible to examine students' ability to analyse the application of physics knowledge in non-standard situations. [6, 11]

Task content is selected according to real-life situations which they might face in everyday life and would be able to deal with them successfully.

4. RESEARCH ANALYSES AND FINDINGS

The problem that is discussed in this article- if a diagnostic tool created in a traditional way for the paper-based environment is applicable in a different environment (in Online/ Computer-based), whilst maintaining its theoretical and technical specific features and ensuring consistent and comparable measurement of results. The research aims – to justify the adaptation of diagnostic tools created in a traditional environment. The following factors that have impact upon the quality of Online/Computer based tests be grouped into three essential groups (First-level cognitive activity tasks, Second-level cognitive activity tasks and Third-level cognitive activity tasks).

This /research analyses the achievements of students on the basis of their skills and cognitive level. After compiling and processing results it was analysed. The average score (Online/Computer-based) in the country are 18, 09 points (2015/2016) and 18.67 (2016/2017) from total 35 points. Diagnostic test in physics for Grade 10 in 2015/2016 and 2016/2017 school years (Figure.1 and Figure.2)

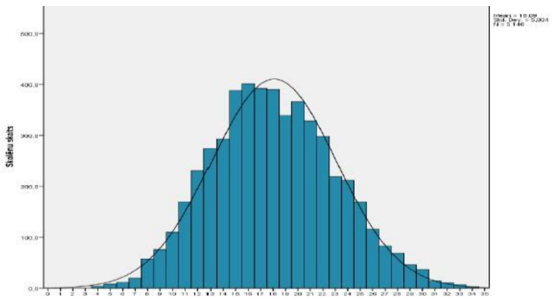

Figure 1: 2015/2016 Point division. The average score in points

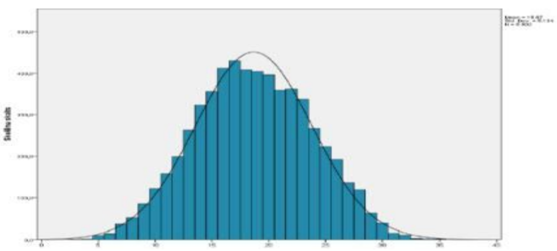

Figure 2: 2016/2017 Point division. The average score in points

Histogram (s. Figure.1 and Figure.2) clearly shows that in Latvia there is a relatively small amount of students who have acquired a very high or outstanding achievement level in physics. Starting secondary school / gymnasium 2016/2017 only 254 (4.38%) students could accomplish more than 77% of diagnostic test tasks, which means they had responded correctly Histogram (s. Figure.1 and Figure.2) clearly shows that in Latvia there is a relatively small amount of students who have acquired a very high or outstanding achievement level in physics. Starting secondary school / gymnasium 2016/2017 only 254 (4.38%) students could accomplish more than 77% of diagnostic test tasks, which means they had responded correctly to at least 27 test questions. This figure compared with the previous year decreased by almost 1% (in the previous year there were 5.27% of students with such result). 2015/2016 there was 18.09 points). The smallest number of points was 3 points (1 student) and 4 points (3 students). None of the students could fulfill all the tasks correctly. The highest result – 33 points – was scored by 3 students and 34 points, scored by 2 students. The average task performance in the country was about 53.35%. Comparing students' results in the physics diagnostic test (Online/Computer-based) for 2016/2017 within specific curriculum aspects, it can be concluded that the range differs by not more than 20%. The summary of acquirement of physics curriculum aspects at a basic educational level included in the diagnostic test is in Figure. 3, while Table 4 compares diagnostic test results in physics for grade 10 results of this and the previous year:

Table 4: The Results of Acquirement of Different Curriculum Aspects

No	***Curriculum aspects***	***2016/2017***	***2015/2016***
1	Physical terms and units of measurement	52,74%	56,70%
2	Understanding and usage of physics concepts	41,30%	
3	Interconnections of the physical quantities	57,20%	48,04%
4	Explaining an experiment	49,48%	49,48%
5	Understanding processes	60,33%	43,66%
6	Graphical representation of physical process, circuit diagrams vectors Understanding and use of models	55,52%	52,68%
7	Scientific enquiry	56,87%	58,10%
Total		53,35%	51,69%

The results of acquirement of different curriculum aspects

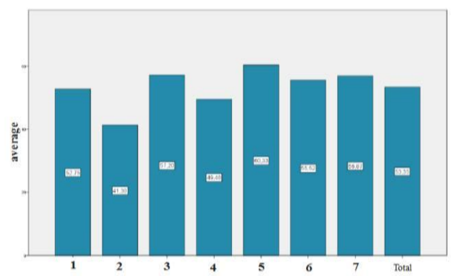

Figure 3: The results of acquirement of different curriculum aspects

Analysis of students' achievements based on their cognitive activity levels (s. Table 4) shows that the students have managed tasks better in general, which require the ability to apply knowledge and skills in various familiar situation. The range of task results in the 2016/2017 diagnostic test is slightly higher than the previous year: the lowest score is 16.10% (task 16), while the highest - 91.96% (task 33) (s. Figure 3). In 2015/2016 the lowest score was 18.93% and the highest - 82, 99%). Diagnostic test results do not show convincing correlation to speak about the insufficient acquirement of a particular skill or particular curriculum aspect.

In each curriculum aspect the average task performance score is 40-60% which is higher than the average score in the whole diagnostic test (s. Figure. 4). It means that the proportion of relatively easy and difficult tasks in the (Online/Computer-based) National diagnostic test was about the same.

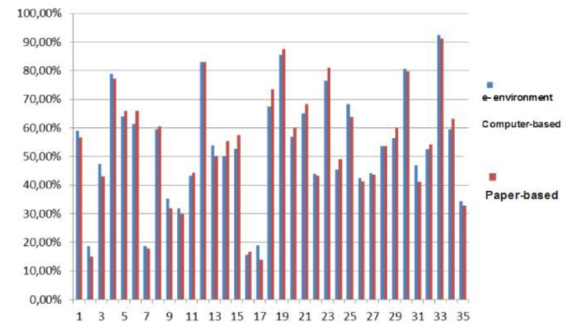

Figure 4: Comparison of the data collected assessing paper-based test and (online/ Computer-based test)

At least 70% of students provided correct answers to the relatively easy six tasks 4, 12, 19, 23, 30 and 33 which is 17% of all tasks in the test. Students' performance in four tasks 2, 7, 16 and 17 or in 11% of all tasks was lower than 20%, while the three tasks 9, 10 and 35 were successfully managed by 30 - 40% of students. This shows that students cope quite successfully with standard/typical tasks (in Computer-based test), but have difficulties with tasks which require high thinking skills and logical reasoning. 22 tasks or nearly 63% of all diagnostic test tasks are classified as tasks at medium difficulty level, because 40 - 70% of students managed them successfully.

For comparison: in Paper-based test there were 8 tasks which were answered correctly by more than 70% of students; there were 2 tasks which were fulfilled correctly by less than 20% of students. At the same time, 18 tasks were classified as medium difficult tasks, because 40-70% of students answered them correctly.

This allows us to conclude that the proportion of easy, medium and difficult tasks in Online /Computer-based and Paper-based was similar. The result of level 2 (s. Figure. 5) cognitive activity tasks - 54% - has remained almost the same as in the diagnostic test of the previous year. Data analysis showed that 1% of students did not mark any option or marked more than one option in one of eight level 1 cognitive tasks and in twelve out of 23 level 2 tasks. This suggests that these students were not were confident in their responses, and/or upon completion of the test, forgot to return to the unfinished tasks (in Computer-based test).

The result of conducting level 1 cognitive activity task compared to the Paper-based test on average decreased by 10%. Many students were not able to recall and recognize acquired physics language elements - concepts and units of measurement, or to recognize their symbols etc., or they had misconceptions about them.

On formal evaluation it is clear that there was an improvement in the results of level 3 cognitive tasks when students had to deal with an unfamiliar situation, analyse it and provide an answer to the question formulated in the task. We need to admit that using multiple-choice tasks makes it impossible to objectively evaluate students' creative action skills. According to many experts' opinions not all tasks correspond with cognitive activity level. The performance percentage of complex physical content tasks 13, 32 and 35 is 52%, 53% and 34%. In addition, responding to 3 from 4 tasks of cognitive level 3 tasks, 1% of students had not marked any option or in contrast marked several options (in Paper-based test)

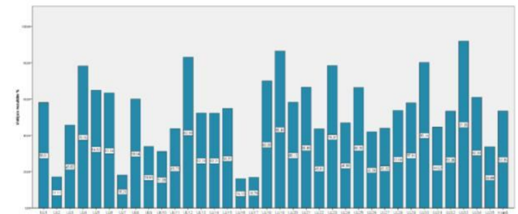

Figure 5. Content of diagnostic test and detailed analysis of each task

A large proportion of students relied on their own every-day-life experience when conducting some tasks, which led them to wrong conclusions. It is also possible that many students did not read the task instructions attentively enough and did not properly analyse every option in the multiple-choice tasks. It should be emphasized that a number of researchers [11] admit that the main difficulties in teaching natural science, and particularly physics, lies in a conceptual change of students' perception and understanding of the world, meaning, to promote a scientific vision of the world, which is often different from students' daily-life experience or naive perceptions. For example, physics education researcher Brown [11] recommends using examples that are closer to students' perception and understanding in the learning process, so that by analysing such examples students will come to new and more general conclusions, which will replace their misconceptions.

[6] The researcher points to the need, when analysing a new situation, to encourage students to find similarities with the known and familiar "transfer" this knowledge to the new situation and to create visual models.

Slightly more than 55% of students have the ability to use physical models (vectors, schemes) and to analyse the information contained in the graphs.

Slightly more than 55% of students have the ability to use physical models (vectors, schemes) and to analyse the information contained in the graphs. Slightly more than 55% of students have the ability to use physical models (vectors, schemes) and to analyse the information contained in the graphs.

Only 44% of students could extract the required information from the heating graph [6], and on description of the situation 58% of students were able to recognize the time-distance graph. 44% of students could comprehend the conductor parallel connection and electrical fuse models. [9] It was already mentioned that test task 30 to a great extent examined students' attentiveness and their skill to compare numerical values of planetary diameters and its representation in the model; the performance of this task showed a high score - 80%.

Students' scientific enquiry skills acquisition level is 57%. Compared to the previous year, it has slightly decreased. Each of the tasks (which was supposed to examine these skills) contained both textual and visual information, which students had to read carefully, analyse and evaluate given options. It is possible that the students were already tired, and like in task 31 which examined students' skills to choose necessary measuring instruments, the correct answer was given by 44% of students, while 42% of students made an inadvertent error, choosing the option D, which along with the necessary measuring instruments contained one redundant measuring instrument. In task 32, 53% of students were able to choose the right hypothesis, which was proved by data obtained in the experiment and presented in the graph. This confirms that more than half of students demonstrated in this task the skill to not only collect the information from the graph, but also come to conclusions.

It is clearly discernible in 22 task, or 80% of incorrect option choices and it shows that a large proportion of students, mostly with low achievements level, conducted the tasks without having read them properly and did not analyse all the options offered in the tasks.

5. CONCLUSION

Research and analysis carried out for the purposes justify the claim is topical for the general public and useful for knowledge sharing and promotion of trilateral cooperation (NC for Education, Latvian Secondary Schools, and Teachers). Since the schools volunteered to conduct the ND tests, the results only partially reflect 10th grade students' achievements in physics. However, the amount of participants in total was large enough (N = 10 948) to draw objective conclusions about students' knowledge and skills, as well as about their weak and strong sides on standard requirements acquisition in physics at a basic education level. Comparing the statistical data of the results of this and the previous school year and some tasks examining certain curric-ulum aspects, come to a conclusion about the improvement of students' learning achievements. That also shows:

1)That there is no significant difference between the results gained from Online/Computer-based test and paper-based National diagnostic tests.

2) Although the tasks' performances in an Online/Computer-based and the traditional way differed (Online test: the lowest result is 18.93%, and the highest - 85.27%; paper based test: the lowest result is 16.71%, and the highest - 94.31%) no compelling correlations were observed which would prove that examining a particular skill or curriculum aspect would have had essentially different results. Assessing students' achievements on their cognitive activity level or any other criteria there is no statistically significant difference.

REFERENCES

[1] A.Kapenieks, B.Zuga, I.Slaidins, I.Vanaga, Dz.Tomsons, A.Kukuka, A.Ulmane-Ozoliņa, V.Rozenbergs, A. Škute, "University Cooperation in Knowledge Society: Modelling in Latvia".Ventspils: proceedings of in-ternational IT conference, 2003.

[2] (Geske, A., Grīnfelds, A., Kangro, A., Kiseļova, R, (2013b) Latvija OECD Starptautiskajā skolēnu novērtēšanas programmā 2012 – pirmie rezultāti un secinājumi. Rīga, LU Pedagoģijas, psiholoģijas un mākslas fakultātes Izglītības pētniecības institūts, 74 pp.

[3] LR Saeima (2010). Ilgtspējīgas attīstības stratēģija līdz 2030. gdam..

[4] Eurydice ziņojums, Galveno kompetenču pilnveide Eiropas skolās: rīcībpolitikas uzdevumi un iespējas. Internetā http://www.viaa.gov.lv/files/news/13703/viaa_galveno_kompetenchu_pi lveide_13_10_2014_fin.pdf.

[5] Guntars Catlaks Kompetenču pieejas ieviešana vispārējā 2015.(ziņojums).Inter-netā.http://www.daba.gov.lv/upload/File/Prezentacijas/150911_NVO_forums_Kompetences_Guntars_Catlaks.pdf

[6] A.Cābelis , L. Jonane, U.Dzerve Diagnosticējošais darbs fizikā re-zultātu analīze un ieteikumi Metodiskais materiāls© Valsts izglītības satura centrs (**National Centre for Education)** 2016 pp.40.-41

[7] EC (2013) Final Report Summary - ANALOGIES - KAPON (Analogi-cal Reasoning and Conceptual Change in Physics Education).

[8] Jānis Kapenieks. Izglītības darbības pētījums estudiju vidē. Promocijas darbs . Rīga [RTU] 2013., 227.lpp. Internetā https://www.liepu.lv/uploads/files/KapenieksJ_Promocijas%20darbs.pdf.

[9] Latvijas rRepublikas tiesību akti. Internetā http://likumi.lv/doc.php?id=20243; http://likumi.lv/ta/id/282340-grozijumi-izglitibas-likuma; http://likumi.lv/doc.php?id=257229#piel27&pd=1

[10] Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions Opening up Education: Innovative teaching and learning for all through new Technologies and Open Educational Resources { 654 final, Brisele, 25.09.2013.) Internetā. http://ec.europa.eu/transparency/regdoc/rep/1/2013/EN/1-2013-654-EN-F1-1.Pdf

[11] A. Geske, A.Grīnfelds Izglītības pētniecība Rīga : LU Akadēmiskais apgāds, 2006. - 261 lpp. ISBN 9984802035(HP)

[12] Eurydice ziņojums 2012. Galveno kompetenču pilnveide Eiropas skolās: rīcībpolitikas uzdevumi un iespējas EACEA — Eurydice nodaļa un Rīcībpolitikas atbalsta nodaļa). ISBN latviešu valodas izdevumam: 978-92-9201-618-0 DOI latviešu valodas izdevumam: 2797/98632 http://eacea.ec.europa.eu/education/eurydice/

[13] 19. EPALE ziņojums 30/10/2015. Izglītības vide un apraksta kritēriji. Internetā. https://ec.europa.eu/epale/lv

[14] Richland,L.E,& Simms, N. (2015). Analogy, higher order thinking, and education. *WIREs Cognitive Science, Volume 6, issue 2, 177-192.* https://doi.org/10.1002/wcs.1336

[15] Betels Dž. (2002).Ievads pārbaudes darbu statistiskā analīzē. Palīgs skolotājiem. IZM, ISAP, Rīga.

[16] European Commission. Europe 2020. International. http://ec.europa.eu/europe2020/europe-020-in-a-nutshell/targets/index_lv.htm, 2011

[17] Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). Education for life and Work. Washington, DC: National Academies Press.

[18] Huberman, M., Bitter, C., Anthony, J., & O'Day, J. (2014). The shape of deeper learning: Strategies, structures, and cultures in Deeper Learning Network High Schools. American Institutes for Research.

[19] Mark A. Colbert, Major, USAF (2001)Statistical analysis of multiple choice testing

[20] Attali, Yigal and Fraenkel, Tamar. —The Point-Biserial as a Discrimina-tion Index for Distractors in Multiple-Choice Items: Deficiencies in Us-age and an Alternative," Journal of Educational Measurement, vol. 37, no. 1, (Spring 2000), 77-86. https://doi.org/10.1111/j.1745-3984.2000.tb01077.x

[21] Ballantyne, Christina. (2000). —Multiple-Choice Tests: Test Scoring and analysis. http://cleo.murdoch.edu.au/evaluations/pubs/mcq/scpre.html.