

COMPARISON OF THE NATIONAL DIAGNOSTIC PAPER-BASED AND ONLINE TESTS IN NATURAL SCIENCE

Loreta Juškaite

Riga Technical University, Latvia

Abstract. *In 2016/2017 Latvian schools were offered the opportunity to choose how to organize the diagnostic test – paper-based or Online. Online 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. 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. What is the difference between the results obtained fulfilling tasks in paper-based and online tests? Are there any differences among different ages when conducting paper-based and online tests? The natural science test was conducted among pupils of the 6th grade (11-13 years old). The aim to examine whether the traditional paper-based tests are applicable in a different environment (in an e-environment Online 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 subjects, which includes using innovative methods in the Latvian educational context.*

Keywords: *Cognitive activity level, Skills, National Diagnostic Test, Online test, Information Communication Technologies (ICT).*

Introduction

A National diagnostic (ND) 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. (Requirements of the Cabinet of Ministers Regulations Nr. 468, 2014) In order to get valid and reliable data the participation of educational institutions from all regions of the country and all types of educational institutions as well is necessary. Should ND tests proceed the traditional way or are there major changes needed? The aim of ND 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. 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? (Kapenieks et al., 2003)

In 2016/2017 Latvian schools were offered the opportunity to choose how to organize the diagnostic test – paper-based or Online. 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. The technical support of the online diagnostic tests provided the portal uzdevumi.lv. 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 ICT to examine students' knowledge. According to the collected data, different types of educational institutions from virtually all regions of the country were represented. In the analysis diagnostic tests performances of 16305 students were used (VISC statistika, 2017).

The research problem – if a diagnostic tool created in a traditional way for the Paper-based environment is applicable in a different environment (in Online), whilst maintaining its theoretical and technical specific features and ensuring consistent and comparable measurement of results.

The research objectives– to justify the adaptation of diagnostic tools created in a traditional environment in the Latvian educational context.

Research methods:

- 1) analysis of scientific literature;
- 2) descriptive statistics and dependency analysis were used to process the data.

The analysis of the research data was carried out using Classical Test Theory (CTT) and Test Analysis Program ITEMAN™ for Windows Version 3.50. The charts and tables were created using MS Excel.

The situation in the Latvian Education System and Education Development Guidelines

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. This should be viewed as a negative factor for the further development of the economy. In the information report of the Ministry 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 increasing 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 %). (Requirements of the Cabinet of Ministers Regulations Nr. 281, 2013) 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). (Eurydice ziņojums, 2014) 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 (LR Saeima, 2010). 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 (Geske & Grīnfelds, 2006). They can consistently distinguish scientific and non-scientific questions, explain the purposes of enquiry, and control relevant variables in a given scientific enquiry or any experimental design of their own. They can transform data representations, interpret complex data and demonstrate an ability to make appropriate judgments about the reliability and accuracy of any scientific claims. At this level students consistently demonstrate advanced scientific thinking and reasoning requiring the use of models and abstract ideas and use such reasoning in unfamiliar and complex situations. They can develop arguments to critique and evaluate explanations, models, interpretations of data and proposed experimental designs in a range of personal, local and global contexts. (Geske, Grīnfelds, Kangro, & Kiseļova, 2016)

Materials and methods

Classical Test Theory

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 (Attali & Tamar, 2000):

- Test score or observed score;
- True score;
- 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 (Ballantyne, 2000):

- True scores and error scores are uncorrelated;

- The average error score in the population of examinees is zero;
- Error scores on parallel tests are uncorrelated.

Test Analysis Program

Results were obtained analysing the diagnostic tests with ITEMAN Test Analysis Program ITEMAN™ for Windows Version 3.50. The ITEMAN software program analyses multiple-choice questions and can also compute and record test scores. The offers four statistical measures:

- Ease Index (called —Proportion Correct”);
- Discrimination Index;
- Biserial;
- Point Biserial Correlation Coefficients.

ITEMAN requires that the input data file be formatted in ASCII (text-only) files. Most data files produced by optical scanning devices are very close to the format that ITEMAN requires, with the exception of the four lines that must be added at the beginning. 5 These lines contain the control line, the key, number of alternatives, etc. The program can process up to a 750-item test with unlimited number of students (Kehoe, 2005).

The user can also manually generate a data file using the edit menu in ITEMAN, which is similar to Windows' Notepad program. ITEMAN's controls are few in number and very simple to use. ITEMAN produces an output file, score file (if desired) and statistics file (if desired). The output file contains the statistical measures, and displays them not only for each question, but for each alternative as well. ITEMAN is a very user-friendly program partly because it has so few controls and only one output file. The ITEMAN program is specifically designed for item analysis only. ITEMAN is easier and faster to use (ITEMAN Online Manual, 2005).

Description of the diagnostic test

Diagnostic tests contain tasks to examine knowledge and skills, the acquisition of which is required for the successful continuation of education. The paper based test is composed of two variants. Online diagnostic test's tasks are mixed. The testing time of both formats – 40 minutes (VISC, 2017).

The ND test is composed of one variant and consists of 24 multiple-choice tasks and 6 open questions. Students have to choose one correct answer from four options (1- 24 tasks). Should answer the question in writing (25- 30) Task content is selected according to a certain level of the mandatory learning content in natural science. In order to assess students' achievements, the test contains tasks with different difficulty and cognitive levels, the content of which relatively covers all aspects contained in the education curriculum in natural science (s. Table 1) (VISC uzdevumi, 2017).

First-level cognitive activity tasks correspond to one step of the operation or procedure. For example, the student remembers the concepts, definitions, units, interconnections or principles, recognizes characteristics, devices or processes. Second-level cognitive activity tasks focus on phenomena and understanding of processes usage. Whereas third-level cognitive activity tasks make it possible to examine students' ability to analyse the application of natural science knowledge in non-standard situations. 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.

Table 1 The proportion of topics in diagnostic test

Curriculum topics	Cognitive levels %			Total
	<i>Memorization and understanding</i>	<i>Usage of knowledge and skills</i>	<i>Analysis and productive activity</i>	
Organisms and life processes.	11	12	3	26
Earth and its place in the Universe.	5	6	3	14
Substances and materials.	10	11	4	25
Physical processes.	7	13	3	23
Environment.	6	4	2	12
Total %	39	46	15	100

Research analyses and findings

The amount of participants in total was large enough (s. Table2)

Table 2 The table captions should be placed above the table

	<i>The number of students</i>	<i>Percentage</i>
Online tests	5794	35,5 %
Paper-based tests	10511	64 %
Total	16305	

The following factors that have impact upon the quality of Online tests be grouped into three essential groups:

- First-level cognitive activity tasks;
- Second-level cognitive activity tasks;
- 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.

A diagram (s. Fig. 1 and Fig. 2) clearly shows none of the students could fulfil all the tasks correctly. The average task performance in the country was about 64.91 %. Analysis of students' achievements based on their cognitive activity levels (s. Table 1) shows that the students have managed tasks better in general, which require the ability to apply knowledge and skills in various familiar situations.

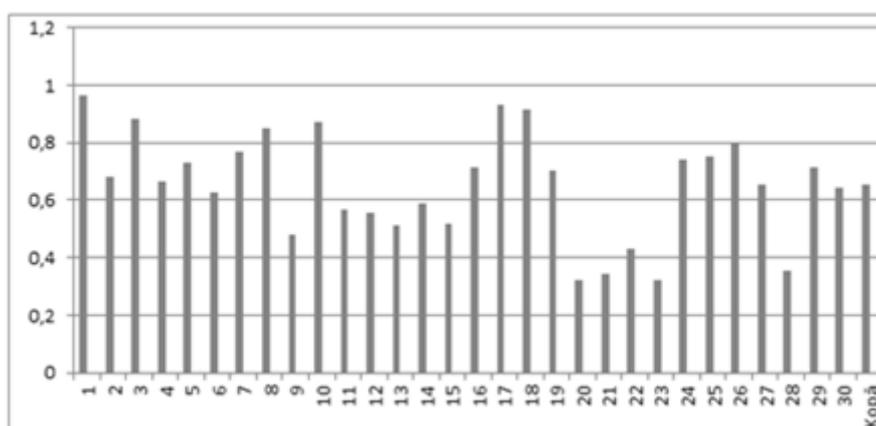


Figure 1. The results (Paper- based tests)

The range of task results Online tests test is slightly higher than the Paper-based tests (s. Fig. 1 and Fig. 2).

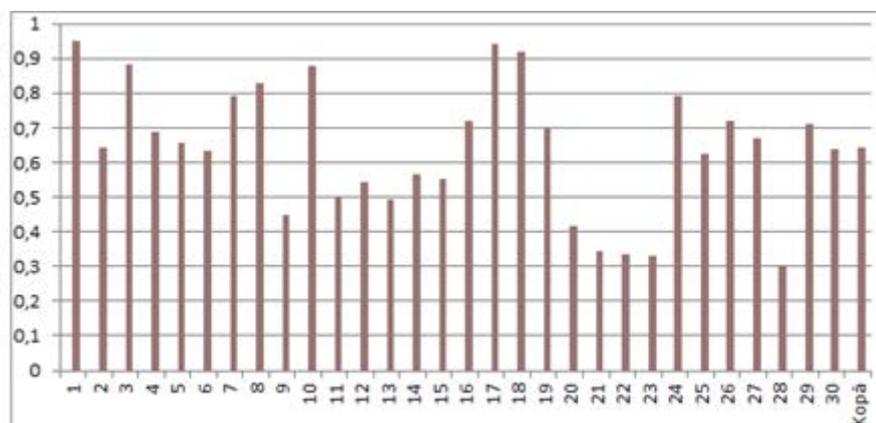


Figure 2. The results (Online tests)

ND 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. Fig. 3). It means that the proportion of relatively easy and difficult tasks in the (Online and Paper - based) ND test was about the same. At least 70 % of students provided correct answers to the relatively easy six tasks 1, 3, 8, 9, 17 and 18 which is 17 % of all

tasks in the test. Students' performance in four tasks 9, 21, 23 and 28 or in 11 % of all tasks was lower than 20 %, while the three tasks 16, 26 and 29 were successfully managed by 30 – 40 % of students. This shows that students cope quite successfully with standard/typical tasks, 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.

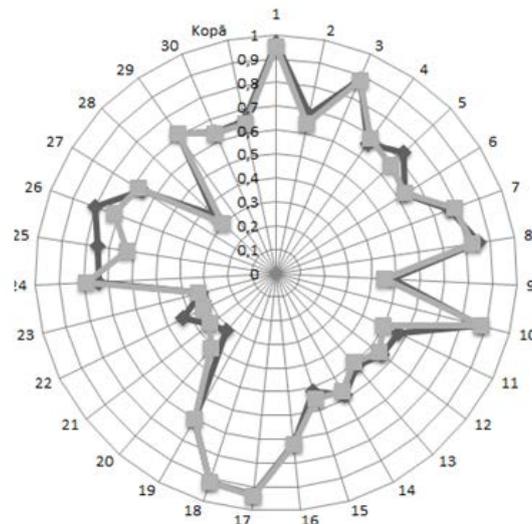


Figure 3. The average results

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 (s. Fig. 3).

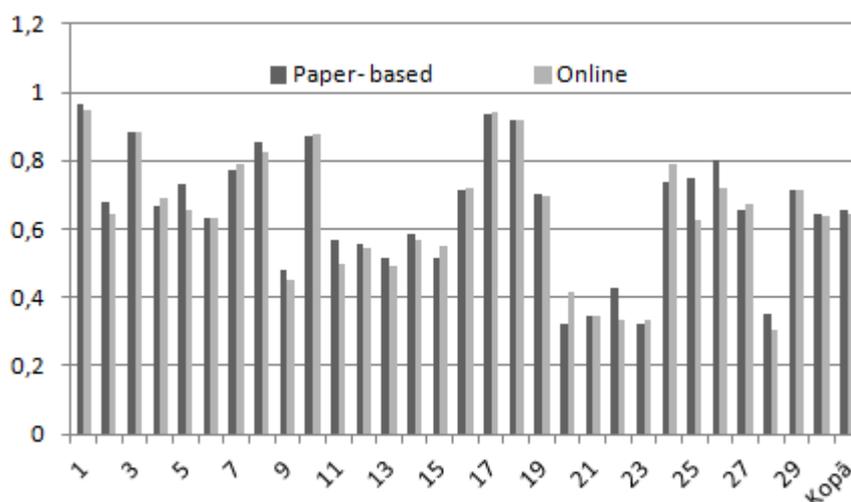


Figure 4. Comparison of the data collected assessing Paper- based tests and Online tests

This allows us to conclude that the proportion of easy, medium and difficult tasks in Online and Paper-based was similar. The result of level 2 cognitive activity tasks – 54 %. This result indicates that almost half of the 6th grade students have insufficient knowledge in the basic education course in natural science. Students are not used to struggling with problems and devising a solution, as well as analysing and coming to the correct answer. Students who conducted the diagnostic test online needed around 32 minutes, although the available test performance time is 40 minutes. 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 2 and 3 level tasks. This suggests that these students were not confident in their responses, and/or upon completion of the test, forgot to return to the unfinished tasks (in online 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 natural science language elements - concepts and units, 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. Author need to admit that using multiple-choice tasks makes it impossible to objectively evaluate students' creative action skills. 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).

A large proportion of students relied on their own everyday-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 (Geske et al., 2016) 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. In this connection, teachers are recommended to try out some unconventional methodological techniques. For example, education researcher Brown (Geske et al., 2016) 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. (Muhammad Anwar, 2013) 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. It should be especially noted that in the performances of a number of tasks it is apparent that

the choice of wrong answer differs less than 2 times. This indicates that a large proportion of students did not properly analyse the task. This tendency was observed in tasks which examined students' scientific enquiry skills to use models and interpret natural science processes. Evaluating the six tasks' performance which contained unusual situations for students which they had to analyse in order to relate the data in the task with the acquired knowledge in natural science – synthesize – and choose the most appropriate option, author can conclude that these higher level thinking skills are possessed by less than 45 % of students. Choosing one of the three incorrect multiple-choice options (distractors), a higher percentage of students often chose answer A or B. It is clearly discernible in 28 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.

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 (National Centre for Education, Latvian Secondary Schools, and Teachers). Since the schools volunteered to conduct the National diagnostic tests, the results only partially reflect 6th grade students' achievements in natural science. ND Online/Computer- based tests innovative IT to examine students' knowledge in the Latvian educational context.

The amount of participants in total was large enough (N = 16305) to draw objective conclusions about students' knowledge and skills, as well as about their weak and strong sides on standard requirements acquisition in natural science at a basic education level. That also shows that there is no significant difference between the results gained from Online test and Paper-based National diagnostic tests. After analysis of the results it can be concluded that students have acquired the most important concepts in natural science - units, research enquiry steps and the ability to reproduce them. There is no significant difference between level 1 cognitive activity task performances Online test and in the traditional Paper-based form.

The main difficulties for students occur in understanding natural science processes, and in analysing and synthesizing a variety of topics using acquired knowledge and skills; this means solving tasks which require the use of acquired knowledge in real life situations and performing higher cognitive activities - analysis and synthesis. There are significant differences in the results comparing Online performances and traditional paper-based performances of the tasks 6, 7, 11, 20, 22 and 24 - on average 8 %. The author can assume that this is due to the students' attitude and the lack of motivation to read the text carefully. In an e-

environment the students had better results in the tasks which included graphical information. Although the tasks' performances in an Online 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.

Recommendations:

That implementation and development of Online testing technologies and pursuit of their quality cannot be an end in itself. (Juškaite, 2017)

Using a blended strategy of national diagnostic development creates more possibilities to avoid mistakes, waste time and material resources and achieve better measuring students' learning achievements results. Automatic test assessment and data transfer to the National Centre for Education. Saving of time and school resources. At the same time Online 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.

References

- Attali, Y., & Tamar, F. (2000). *The Point-Biserial as a Discrimination Index for Distractors in Multiple-Choice Items: Deficiencies in Usage and an Alternative*. *Journal of Educational Measurement*, 37, 77-86.
- Ballantyne, C. (2000). *Multiple-Choice Tests: Test Scoring and Analysis* Retrieved from <http://cleo.murdoch.edu.au/evaluations/pubs/mcq/scpre.html>
- Cābelis, A. (2016). *Metodiskais materiāls*. R., Valsts izglītības satura centrs.
- Eurydice ziņojums (2014). *Galveno kompetenču pilnveide Eiropas skolās: rīcībpolitikas uzdevumi un iespējas*. Retrieved from http://www.viaa.gov.lv/files/news/13703/viaa_galveno_kompetenču_pilnveide_13_10_2014_fin.pdf
- European Commission. ziņojums *Europe 2020*. (2017). Retrieved from http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/targets/index_lv.htm, 2013
- Geske, A., & Grīnfelds, A (2006). *Izglītības pētniecība*. Rīga, LU Akadēmiskais apgād.
- Geske, A., Grīnfelds, A., Kangro, A., & Kiseļova, R, (2016). *Latvija OECD Starptautiskajā skolēnu novērtēšanas programmā 2016 – pirmie rezultāti un secinājumi*. Rīga, LU Pedagoģijas, psiholoģijas un mākslas fakultātes Izglītības pētniecības institūts.
- Juškaite, L. (2017). Analysis of the National diagnostic paper-based and online tests in Physics. *ICCMIT 2017, Warsaw conference materials*, 61- 62.
- Kapenieks, A., Zuga, B., Slaidins, I., Vanaga, I., Tomsons, Dz., Kukuka, A., Ulmane-Ozoliņa, A., Rozenbergs, V., & Škute, A. (2003). University Cooperation in Knowledge Society: Modelling in Latvia. *Ventspils: proceedings of international IT conference*, 7- 8.
- Kehoe, J. (2005). *ITEMAN Online Manual*. Assessment Systems Corporation. *Basic Item Analysis for Multiple-Choice Tests*. Retrieved from <http://www.assess.com>.
- Latvijas Republikas tiesību akti*. Retrieved from <http://likumi.lv/doc.php?id=20243>; <http://likumi.lv/ta/id/282340-grozījumi-izglitibas-likuma>;
- LR Saeima (2010). *Ilgspējīgas attīstības stratēģija līdz 2030. gadam*. Retrieved from <http://likumi.lv/doc.php?id=257229#piel27&pd=1>

Muhammad, A., Rehman, P., & Shaheen, P. (2013). Missing Elements of Computer Science Curricula. *Global Journal of Computer Science and Technology Cloud & Distributed*, 12 (11).

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 (Annex 8) Retrieved from <https://likumi.lv/doc.php?id=268342>

VISC, eksāmeni, info 2017 (2017.01.17.) Retrieved from http://visc.gov.lv/vispizglitiba/eksameni/info_old.shtml

VISC, eksāmeni, statistika 2017 Internetā (2017.01.08.) Retrieved from <http://visc.gov.lv/vispizglitiba/eksameni/statistika/2017/dokumenti/06kl>.

Uzdevumi.lv (2017). *Informācija medijiem, 2017. g. 15. marts* Retrieved from <http://www.uzdevumi.lv>