MONITORING OF STUDENTS' EDUCATIONAL ACHIEVEMENTS:
Experience of Nazarbayev Intellectual Schools in Kazakhstan

40th IAEA Annual Conference
Singapore, May 25-30, 2014

Rustam Abilov*, Nico Dieteren+, Frans Kleintjes+, Gulnara Sadyk* and Daniyar Temirtasov*

** Center for Pedagogical measurements, Nazarbayev Intellectual Schools,(Kazachstan)*
+ Cito Institute for Educational Measurement (the Netherlands)

Abstract

One of the key strategic tasks of the "Autonomous Educational Organisation Nazarbayev Intellectual Schools" (hereinafter referred to as NIS) states "Developing a transparent system for quality assurance". Along with internal classroom assessment, NIS is developing, piloting and implementing a system of External monitoring of students' educational achievements aimed as a formative assessment tool. Progress towards meeting the individual student's goals is measured by comparing expected outcomes stated in the curricula and actual rates of learning. Based on these measurements, teaching will be adjusted as needed. Thus, the student's progression of achievement is monitored and instructional techniques are adjusted to meet the individual student's learning needs. The development has started in 2012 with a student monitoring system for mathematics for secondary education and in the development for languages in secondary education has started in 2013.

Some approaches that are new to Kazakhstani context, are used in development of the Monitoring system. The article describes steps undertaken in the development of Monitoring system and experience gained in the development and the piloting of the system in the field of mathematics. Particular attention is paid to the approaches of setting performance standards, development of level descriptors of students' achievements and the different ways of reporting for the stakeholders.

1 Introduction

Autonomous educational organization "Nazarbayev Intellectual Schools" is a network of Intellectual schools that was initiated by the President of Republic of Kazakhstan in 2008. There are 15 schools that are functioning at this stage. Intellectual schools were initiated with the mission to enhance the intellectual capacity of Kazakhstan through the development and implementation of an innovative, mathematics and science-orientated, trilingual, model school system that integrates the best of Kazakhstani traditions, and that meets international standards of best practice.

NIS schools intend to prepare alumni that are patriots of their country with firm moral values, proficient in Kazakh, Russian and English. Schools aim to prepare students that graduate from prestigious national and international universities and compete successfully in the labor market. They shall actively be involved in the socio-economic and political processes in the country and widely interact and innovate at all levels, to expand the influence of Kazakhstan in the international arena.

The goal and missions set require not only the development of content and methods of education that fulfill expectations of the 21 century demands, but also to develop and

implement a system which will allow to systematically track, analyze, evaluate and predict the students' performance during their school life.

Taking into account an analysis of the international expertise and the NIS experience of the existing monitoring system, the goal was set to develop a system, which allows obtaining reliable information on the performance and progress in students learning in the dynamics for the correction an individual learning profile of each student. In order to fulfill the demands for new monitoring system, Cito, the Institute for Educational Measurement (The Netherlands) was chosen as Strategic partner. Cito has a longstanding tradition and experience in providing student monitoring systems for primary and secondary education in the Netherlands.

The development of the student monitoring system for mathematics started in 2012, at the same time a new curriculum was introduced in NIS. It was believed that the new curriculum would be enforced by the introduction of such an extensive assessment program. The new curriculum was introduced in grades 7 and 11 in school year 2011-2012. Hence the development of the student monitoring system is in pace with the introduction of the curriculum and now in its second year of development.

2 Main features and stages of the student monitoring system for mathematics at NIS in Kazakhstan

2.1 Starting situation

- **Close relation to the Integrated Program of Development**

The monitoring system of educational achievements is based on the contents of the new curriculum, called 'Integrated Program of Development'(hereinafter - IPD) developed by experts of NIS together with experts from the Cambridge International Examinations, University of Cambridge;

- **Based on test matrices**

Items are developed on the basis of test matrices, which consist of assessment objectives, formulated on the basis of the learning objectives in the IPD;

- **Five content domains**

The assessments cover five content domains: Numbers, Algebra, Geometry, Statistics and Mathematical Modeling.

- **Taxonomy of Bloom**

The level of cognitive demand of the items are developed according the levels of cognitive skills in Bloom's taxonomy.

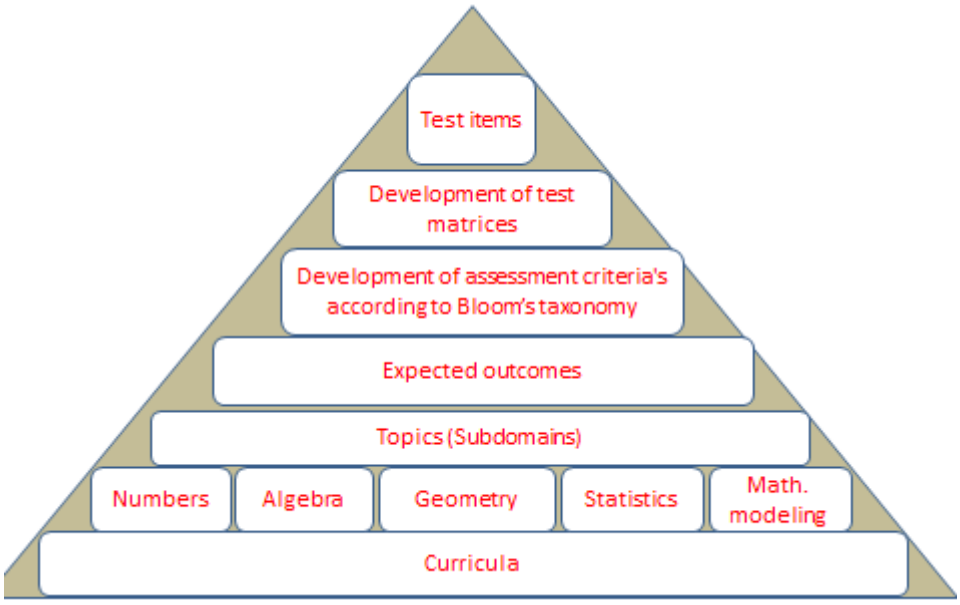
- **30 items per test taking per domain**

To arrive at a reliable and valid measurement each test consists of 30 items, both multiple choice and open ended.

2.2 Some elements of item development in more detail

Item developers are the teachers of Intellectual schools, trained, instructed, supported and partly directed in construction teams by Cito subject experts.

Figure.1. Action algorithm on item development used in monitoring system.



Teachers-developers studied the expected outcomes on domain topics and constructed test matrices, which contained assessment objectives on each topic of the domain, on the basis of which items were developed.
At the initial stage of developing a monitoring system items were developed for the following cognitive levels: knowledge, understanding, application, analysis (Table 1). Later, with the acquired experience in the item development, it is planned to develop items on higher cognitive levels, such as synthesis and evaluation

For each item, there are characteristics that give information about the topic, cognitive level according to Bloom's taxonomy, assessment criteria and the expected level of difficulty.
As noted above, the theoretical basis for the item development is Bloom's taxonomy, which helps to qualitatively formulate assessment objectives. The distribution of items in cognitive levels and examples of verbs are given in Table 1.

Table 1. The share of developed items in mathematics at the cognitive level, Grade 7

Cognitive level	Knowledge	Understanding	Application	Analysis
250 items	5 %	21 %	63 %	11 %

Skills (Examples)	count determine identify outline list	classify compare convert discuss differentiate	calculate construct demonstrate illustrate predict	break differentiate discriminate explore correlate

Screening of developed items for the first monitoring of Intellectual schools students` achievements.

The developed items are screened internally and externally. After the development items are screened within a work group of developers. Considering the item of the colleague, the developer makes a comment on each item, thus receiving a comment of colleagues, if necessary, items are changed, all the work is coordinated by an expert of the Center for Pedagogical Measurements, a branch of NIS. Next, the items are sent to external screening to Cito and, if necessary, on the basis of expertise, items are either deleted or changed. On the basis of this work the item bank is updated.

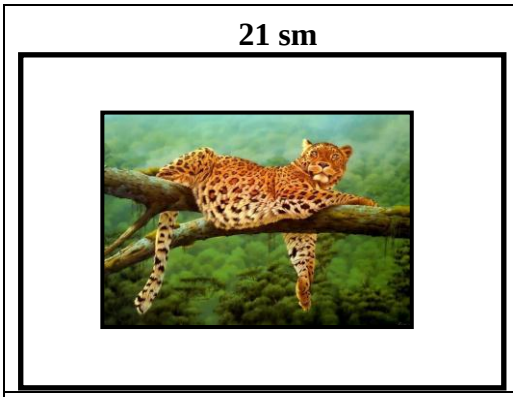
In general, all of the 250 developed items on 7^{th} grades were screened by Cito. After revising and screening the number of items was 246.The number of developed items for 11^{th} grades were 220, 46 of which have been changed and 24 have been deleted and 28 extra items were developed. After revising and screening the number of item was 224.

2.3 Item specifications and item bank

Developed in cooperation with Cito experts items check students` knowledge, ability to apply its functional literacy. The content of items is based on the specific facts and events from real-life situations, including events taking place in Kazakhstan. The technology of item development corresponds to the international standards as well as in item formulation, selection and arrangement of distractors, also in the test design. The items are administered and managed in item bank system.

Figure 2 Sample of test item and its characteristics:

Find the area (in cm^2) of the edging picture pasted on a sheet in the form of rectangle, if the size of the picture 15 cm × 10 cm.



A 155
B 186
C 336
D 486

Item characteristics:

Identification code	7.3.3.4.4

Domain	Geometry
Topic	The metric relations
Thinking operation (Bloom`s taxonomy)	Analysis
Assessment criteria	To explore the area of the figure
Correct answer	B
Expected difficulty level	Hard

2.4 Administration of student monitoring test items

- **Student monitoring three times per school year**

Monitoring is be conducted three times in the school year - at the beginning (September), in the middle (January) and at the end (April) of the school year, depending on the specifics of the subject; For grade 7 and 11 assessment have been developed for these three administration moments. The development for grades 8 and 12 is currently limited to a January administration.

- **From paper based to computer based**

The test is paper based at the moment, but in future it is envisaged to make a transfer towards computer-based testing. For reasons of efficiency and cost saving, paperbased booklets of the most recent administrations are displayed on computer screens for candidates. For answering the candidates use preprinted optical readable answer sheets. Marking still has to be done manually.

- **Pretesting of items**

Piloting of test items is held in order to obtain information about the quality and to improve a standardized approach in the administration of the test. After screening and revising the developed items, a pre-test is conducted within the target groups of students in NIS schools. Purposes of pre-testing are quality control and improvement of individual items, of composed test versions and of test administration procedures. Some aspects to be checked:

- the difficulty of each item;
- shortcomings of items with the objective for further improvement;
- final test time, which is required for doing the whole test in general;
- analysis of open items answers possibly to clarify the formulation of correct answers or assessment scheme;
- test design, per booklet and overall;
- technical errors in the test booklet (typos, spelling and punctuation flaws, etc.);
- shortcomings in the instructions for test administration.

2.5 Analysis of test results

- **Use of Item Response Theory**

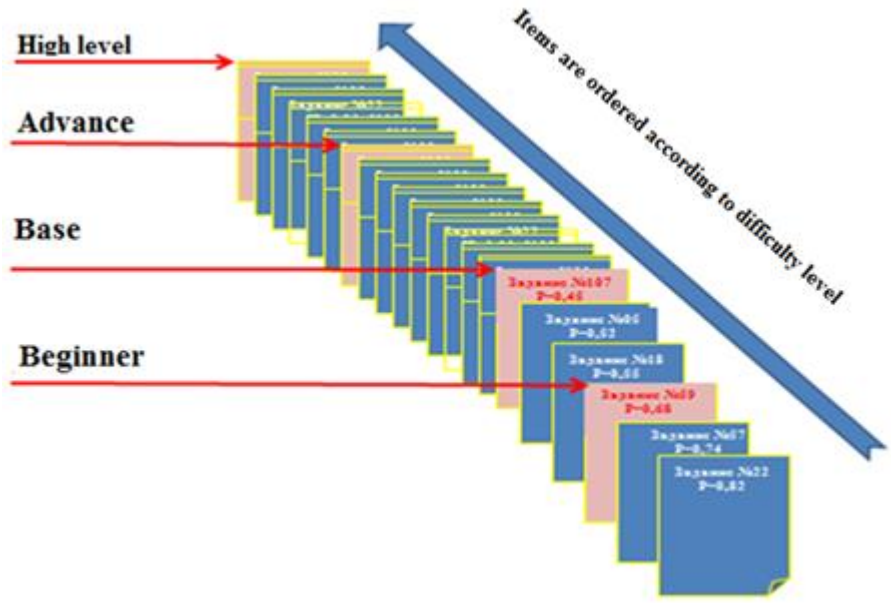
The monitoring results are analyzed using item response modeling. Classical Test Theory (TIAplus) is used to evaluate the test taking and a special item response model, the One Parameter Logistic Model (OPLM) is used to place the achievement of each student onto **one** ability scale per content domain, that will help to follow the student performance path during the entire period of learning in the Intellectual schools.
The same item response model is also used to determine the main item characteristics: difficulty and discrimination.

- **Setting Standards.**

Performance standards per domain and administration moment are set using the ordering of the items found in the item response model. This step is important and necessary as the standards are the tool which help to determine individual students' ability at each administration moment and also progress of this ability during the study period.

For this monitoring system the "Bookmarks" method is used.

Figure 3. Scheme of the "Bookmarks" method



C**ategories.**

Teachers-item developers together with Cito experts allocated 4 levels of assimilation of each domain of the curriculum in mathematics in accordance with the expected outcomes:

1. **Beginner level** – student has general concept of mathematical terminology. Has insufficient level of practical skills.
2. **Base level -** the student is able to give a meaningful definition of terms, concepts, know how to use the language of mathematics, laws, terms and concepts, know how to choose and use appropriate mathematical knowledge, skills, knows how to make consistent conclusions.
3. **Advanced level** – student knows the theoretical material, able to apply knowledge and skills in solving problems in non-simple situations, can prove not simple statements.
4. **High level** – Student is fluent in mathematical language and mathematical apparatus, able to interpret and convert nonstandard problem in a more intuitive and accessible level, establishes the connection between domains.

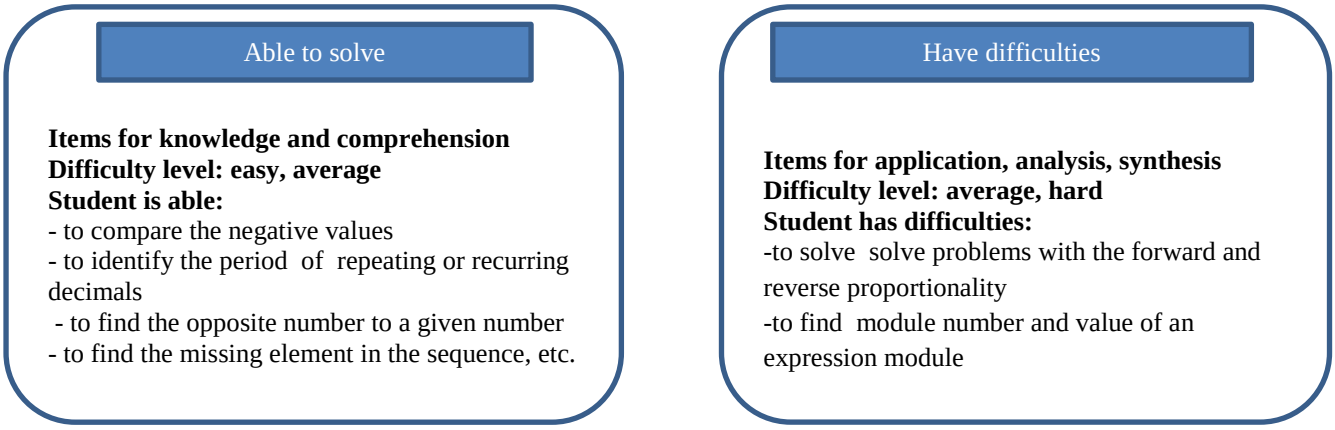
Level descriptors.

Each of indicated levels has detailed description in the context of monitoring i.e. what skills, abilities should student have while conducting a particular monitoring, and what skills and abilities student should repeat. This is called a **level descriptor**.

Below there is an example of a base level descriptor for the mastery of mathematical content of the curriculum in the domain "Numbers":

Description: *student is able to give a meaningful definition of terms, concepts, able to use the language of mathematics, laws, terms and concepts, able to choose and use appropriate mathematical knowledge, skills, knows how to make consistent conclusions*.

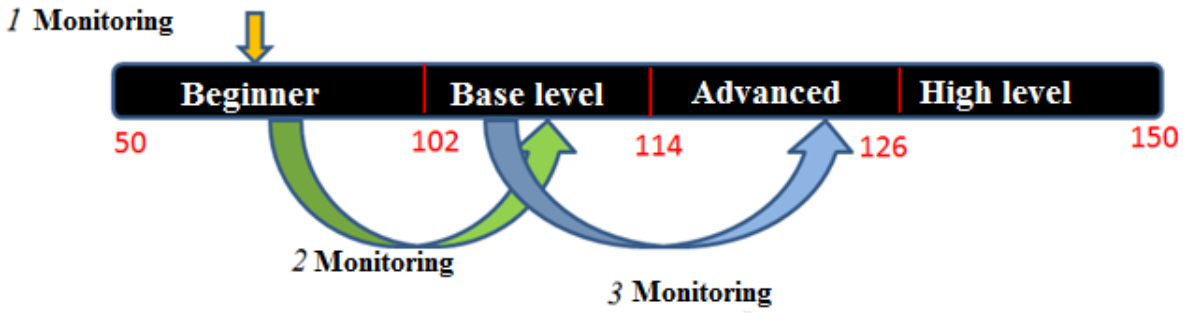
Figure 4. Examples of level descriptors



2.6 Reporting on test results and diagnostic information

The results of monitoring of students' achievements are presented in the form of reports for different categories of stakeholders: for AEO, the school administration and teachers, parents and students.
As it was mentioned above, the system allows tracking individual progress of each student. Change of score in the next monitoring and moving or not moving on the one and the same ability scale from one level to another is a confirmation of the presence or absence of progress in learning a particular domain of the curriculum. Categories belong to '2 Monitoring. Categories belonging to '3. Monitoring' are not shown. The categories are expected to move along the ability scale each administration moment. So what is 'Base level' for 2. Monitoring may be about beginner in one of the following administration moments. Research is still ongoing to determine the actual shifts.

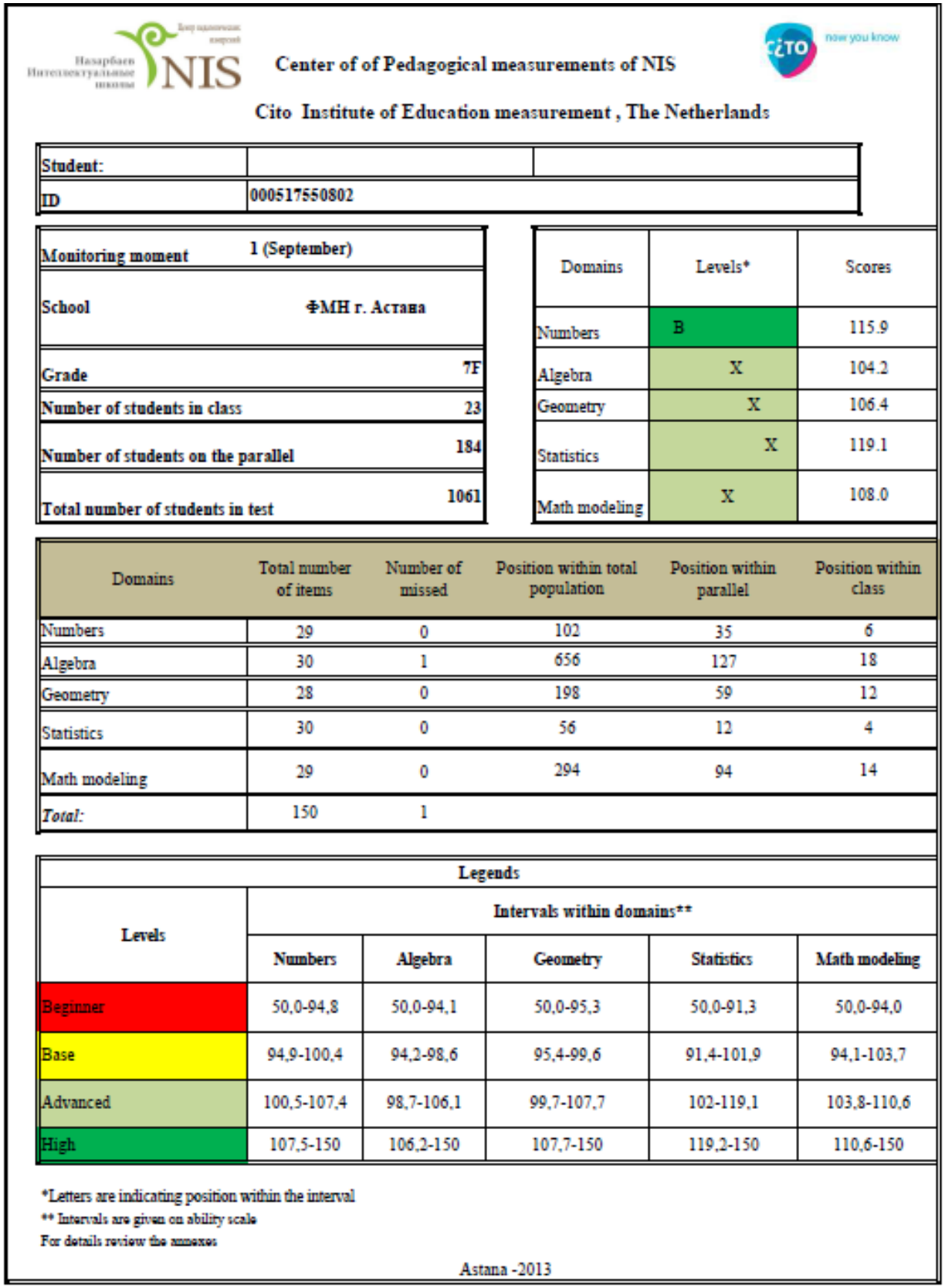
Figure 5. Example of an ability scale (equated scores are used)



An individual student report is developed on the basis of the analysis as described above. In figure 6 an example of an individual student report after one administration is shown. The

ability score of a student, position in the performance category an ranking is presented..

Figure 6. Example of individual student report after first administration moment in grade 7



Center of of Pedagogical measurements of NIS

Cito Institute of Education measurement , The Netherlands

Student:		
ID	**000517550802**	

Monitoring moment	**1 (September)**
School	**ФМН г. Астана**
Grade	**7F**
Number of students in class	**23**
Number of students on the parallel	**184**
Total number of students in test	**1061**

Domains	Levels*	Scores
Numbers	B	115.9
Algebra	X	104.2
Geometry	X	106.4
Statistics	X	119.1
Math modeling	X	108.0

Domains	Total number of items	Number of missed	Position within total population	Position within parallel	Position within class
Numbers	29	0	102	35	6
Algebra	30	1	656	127	18
Geometry	28	0	198	59	12
Statistics	30	0	56	12	4
Math modeling	29	0	294	94	14
Total:	150	1			

Legends

Levels	Intervals within domains**				
	Numbers	**Algebra**	**Geometry**	**Statistics**	**Math modeling**
Beginner	50,0-94,8	50,0-94,1	50,0-95,3	50,0-91,3	50,0-94,0
Base	94,9-100,4	94,2-98,6	95,4-99,6	91,4-101,9	94,1-103,7
Advanced	100,5-107,4	98,7-106,1	99,7-107,7	102-119,1	103,8-110,6
High	107,5-150	106,2-150	107,7-150	119,2-150	110,6-150

*Letters are indicating position within the interval

** Intervals are given on ability scale

For details review the annexes

Astana -2013

Reporting the student progress, actual one of the main features in student monitoring is still being developed. The first drafts have been send to schools for evaluation. Along with a report as shown in Figure 6, teachers and students receive more detailed reports, which allows them to work on the topics or assessment criteria's where student scored low

3 Conclusion

The quality assurance of education in schools is currently unthinkable without tools and technologies that allow to obtain systematically reliable information about the status of the educational process. In pedagogical literature about monitoring and formative assessment of recent year's we can see a gradual shift from assessment for control (checking) to assessment conducive to learning (assessment for support). In developing our monitoring system we try to bring together the diagnostics and educational practice.
Much has been achieved, but may parts of the student monitoring system still has to be developed ánd validated. This process will take considerable time and effort.

List of References

Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc. (!Update of 2009????)

Downing, S.M. &Haladyna, T.M. eds. (2006). Handbook of test development. New Jersey: Lawrence Erlbaum Associates, Inc.

Barton, P. E. (2002). *Staying on course in education reform*. Princeton, NJ: Statistics &

Research Division, Policy Information Center, Educational Testing Service.

EO Nazarbayev Intellectual Schools (2013). Development Strategy of AEO Nazarbayev Intellectual schools until 2010. Retrieved from: http://www.nis.edu.kz/ru/about/str-doc/

Verhelst, N.D., & Glas, C.A.W. (1995) The one parameter logistic model. In: G.H. Fischer & I.W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications* (pp. 215-239). New York: Springer.

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