SUBJECT KNOWLEDGE ASSESSMENT IN THE SELECTION OF STUDENTS TO GRADE 7 OF NAZARBAYEV INTELLECTUAL SCHOOLS

39th IAEA Annual Conference
Tel-Aviv, Israel, October 20-25, 2013

Olga Mozhayeva (NIS, Kazakhstan)
Tynybayeva Madina (NIS, Kazakhstan)
Gulnara Sadyk (NIS, Kazakhstan)
Henk A. Moelands (Cito, The Netherlands)
Frans Kleintjes (Cito, The Netherlands)

Abstract
A new curriculum and assessment aimed to foster 21st century skills are being developed and implemented at Nazarbayev Intellectual schools (NIS). The major of the curriculum is Mathematics and Sciences. Besides a new curriculum with a focus on Mathematics and Science, a trilingual policy will be implemented as well. This would mean that the alumni by finishing NIS are proficient in Kazakh, Russian and English. In order to accept students who would cope with the curriculum and assessment a new system of selecting students has been developed together with NIS strategic partner – Cito, Institute for Educational Measurements, the Netherlands. Taken the mission and major of NIS into account, four subjects have been defined to be included into the selection procedure: Mathematics, Kazakh, Russian and English. Absolutely new for Kazakhstani context approaches to test development, (pre)test administration and (pre)test result analysis have been applied during the process of developing a new selection system. With the support of Cito experts, NIS item developers (teachers) developed an itembank In September 2012 the developed items were pretested among NIS and common schools students. The results of the pilot have been very informative in setting cut off scores decisions and test procedures. In June 2013 a first live administration of the test took place. Analyses showed that the test worked well. At the conference we will present the process of developing the test and the outcomes of the first live administration. Also lessons learned will be discussed and recommendations will be made.

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 labour 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 of student selection that would identify students, that are capable of studying and succeeding in Intellectual schools, further education and professional life.

The system of selection in Intellectual schools, used prior to June 2013 was characterized as bound to the content knowledge, not fully standardized and had low discriminative ability and lack the rationale behind pre-determined cut-off score. In contrast, the system had to

focus not only on content knowledge of the students but also on their ability to apply the acquired knowledge. Moreover, as a high stakes test, it should meet the demands of test and test results security, should discriminate well between able and less able students and should be statistically validated (grounded).

In order to fulfill the demands for new selection system, Cito, the Institute of Educational Measurements (The Netherlands) and Johns Hopkins University Center for Talented Youth (USA) were chosen as Strategic partners. Particularly, The Center for talented Youth is well known for its significant work in the area of identifying and developing the talents of the most advanced learners worldwide. The CTY Ability test which consist of two main parts, Quantitative reasoning ability assessment and Spatial reasoning ability assessment is administered at the second day of Selection. Nevertheless, at this conference we will be talking about the subject tests which have been developed, piloted and implemented by the joint efforts of AEO Nazarbayev Intellectual schools and Cito, The Institute of Educational Measurements.

STRUCTURE OF THE SUBJECT TESTS

Target group
The candidates (11-13 year olds) applying for 7th Grade of Intellectual schools are being selected at each of the 13 cities. Intellectual schools were established since 2008. The number of applicants is increasing each year according to the region. Obviously, the cities of highly populated parts of the Kazakhstan have more applicants. There are 15 functioning Intellectual schools at this stage across the country. However, the vacant places available at each school each year are predefined and are limited. Given the number of available places only 14% - in average - of these candidates can be allowed to start in grade 7 next school year. This percentage varies considerably between different schools and between candidates that apply for classes with Kazakh and Russian as language of instruction.

Structure of the Selection and the Subject test
New student selection tests are administered in two days: on the first day the subject tests, and on the second day the CTY Ability test. In order to select most qualified students minimal cut off scores have been set for Mathematics and the quantitative part of the Ability test. The underpinning argument that lay behind this decision is in the strategic tasks of the Development Strategy of Intellectual Schools which is stated as "searching for and competitive selection of children, who are capable in learning mathematics and natural sciences" (AEO, 2013); Students are then selected according to the highest scores on total Subject tests. For an overview of how the tests are administered, see Table 1.

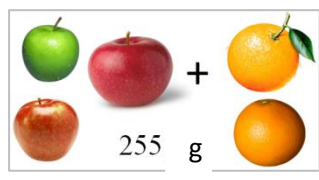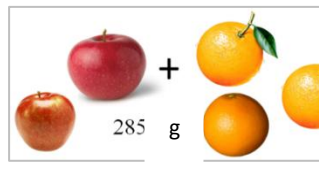Table 1: Overview of tests during day 1 and day 2

Day 1		Day 2	
Subject test (100 items in 180 minutes)		**Ability Test (134 items in 77 minutes)**	
Mathematics 40 items in 60 minutes	Kazakh 20 items in 40 minutes	Quantitative section 60 items in 30 minutes	Spatial section 74 items in 47 minutes
Russian 20 items in 40 minutes	English 20 items in 40 minutes		

The subject tests consist of three subjects. Mathematics, Kazakh as first or second language, Russian as first or second language, and English. The trilingual policy of the AEO requires that potential students can master subjects in all three languages. Students, that are good at

mathematics and science subjects should also be able to read and write and explain orally their ideas and arguments, as well as communicate with others.

Mathematics test items have been developed on the basis of 8 domains (Relations and Proportions; Positive, Negative and Rational Numbers and Absolute value; Linear equations and inequalities; division of numbers; Percentages; Geometry). Items are oriented to assess the ability of students to apply the knowledge (context-related items). 20% of Math items are designed to assess application skills. In accordance with the test specification, the test focus more on higher order skills than is usual in current practice of Kazakhstan. In constructing the test-grid the taxonomy of Bloom (Bloom, 1956) was used. The taxonomy of Bloom describes six abstraction levels that are commonly used when creating test questions and describe lower to higher order skills (Remember, Understand, Apply, Analyze, Evaluate, Create). Mathematics test has been designed to assess technical math skills. To be able to fulfill the Math task, candidates should analyze the given situation and formulate a strategy to solve it. Below an example of a Mathematics item (Figure 1).

Figure 1: Example of Mathematics item of the Selection test

The weight of 3 apples and 2 oranges is 255 g   The
weight of 2 apples and 3 oranges is 285 g

How much do 1 apple and 1 orange weigh?

  39 g
  69 g
108 g
110 g

The language tests consist of Reading Comprehension tests that assess literacy and verbal competences through work with texts. Speaking and Writing production were not used for the selection test, as these skills cannot be tested through multiple choice questions, which was a prerequisite. Based on international research, reading comprehension (in mother tongue and/or in foreign languages) can be considered the best predictive factor for school success in the future.

Kazakh and Russian languages are assessed as mother tongue and as second language. As a result, the assessment objectives that have been used for the first and second language tests are the same, with the provision that the tests for Kazakh and Russian as first languages should be more difficult than the tests for Kazakh and Russian as second languages. This has been realized by varying the difficulty level of the texts and the accompanying questions. This approach meets the curriculum of the common schools of Kazakhstan and the curriculum of Intellectual Schools.

Three groups of assessment objectives have been chosen to create the basis for Language tests matrices, namely:
-        Reading (comprehension);
-        Meaning of words, sentences, paragraphs, texts;
-        Relations and communicative aspects (such as opinions, conclusions, feelings).
An example of an English item is shown in Figure 2.

Figure 2: Example of an English item

You are travelling by car with
your parents and a little sister who is learning to read. You
see the sign:
Whom is this sign for?
A) for people who want to advertise something
B) for people who have lost the way
C) for people who prefer travelling by car

The format of items
Considering the high stakes of the selection test, it was decided to develop multiple choice items. Multiple choice items contribute to avoid unreliable judgments and subjectivity in marking, by eliminating the participation of people in test scoring. Stakeholders will therefore be more convinced that the test procedure is fair. Furthermore, it is much easier to explain stakeholders (students, parents) the scoring process; right or wrong. Moreover, rater-disagreement will not be at stake. And finally, there is an advantage of decreasing time between test administration on the one hand and reporting on test results on the other hand. In future, in case of computer based testing, reports on the results of student selection can be prepared instantaneous, which will be very important. The less time between test administration and reporting the higher the acceptability of the test among its stakeholders will be.

The number of items
A single test for mathematics consists of 40 items, tests for language subjects consist of 20 items each. Aspects that have been taken in consideration in deciding the number of items are the purpose of the test, time available and the age of candidates. The number of items of Mathematics was agreed to be two times more than for each language, due to the high importance of mathematics in learning of science subjects.

Testing time
Testing time for each Mathematics item was set as 1,5 minutes which makes 60 minutes in total for the whole mathematics test. For Language tests, it was decided to allocate more time, as the tests consist of texts, which require extra time for reading. The time allocated for each subject is strictly fixed. This helps to have fair comparison of student results. The amount of time allocated for the completion of one item proved to be reasonable according to the results of piloting the Selection test and procedures. There was no evidence of shortage of time for completion of Mathematics or Language test items, which was also confirmed during live testing in June.

Test versions
In any high stakes test that is meant to be administered more than once, it is important to consider the risk of item exposure. To mitigate this risk, different version of the selection test

can be used. These versions consist of different items drawn from a larger item bank. The different versions overlap so that they can be equated using item-response theory.

Pretest
In order to collect data about the quality of the items and the introduction of new improvements in the procedure, the selection test items and the testing procedures have been pretested. To be able to make reliable inferences about the quality of an item it was decided to have at least 300 observations per item. 666 Intellectual School students and 1114 seventh grade students from common state schools participated in the pre-test.

The pretest design allowed for 240 math items and 120 items per other subject to be pretested. These items were grouped into 24 different booklets in such a way that each booklet was consistent with the test-grid and had a sufficient amount of overlap with other booklets.

As a result of the pretest, some adjustments have been made in order to improve the items and its procedure:
•   categorization of the difficulty of an item - hard/medium/easy - was changed;
•   the wording of some items that performed poorly was changed;
•   the classification of an item according to domain of the subject or assessment objective was changed;
•   item was removed from the item bank, in case it proved to be inappropriate;

Student results according to pretest
The test score distributions for Intellectual schools and Common schools students on the pre-test, expressed as a percentage correct on the whole set of items for the given test, are shown below. In each figure two separate histograms are drawn, a gray one for common school students and a blue one for Intellectual school students. All common school students made both tests for the second language. In the graphs only the results for the tests that actually were the students' second language are shown.
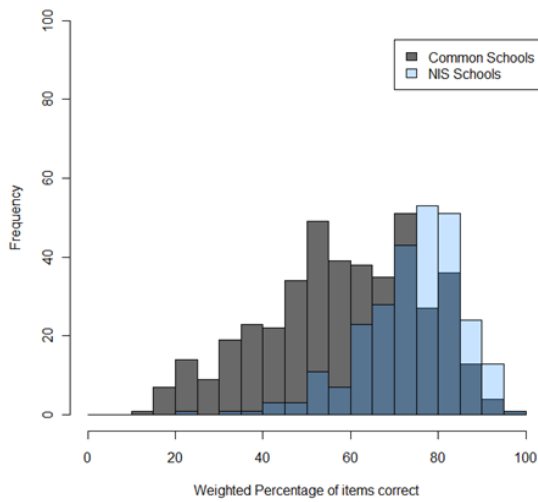As can be seen there is some overlap between score distributions for the two samples but Intellectual school students generally score higher on all subjects compared to common school students.
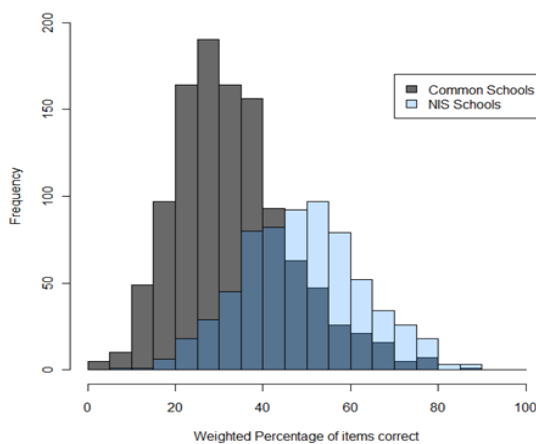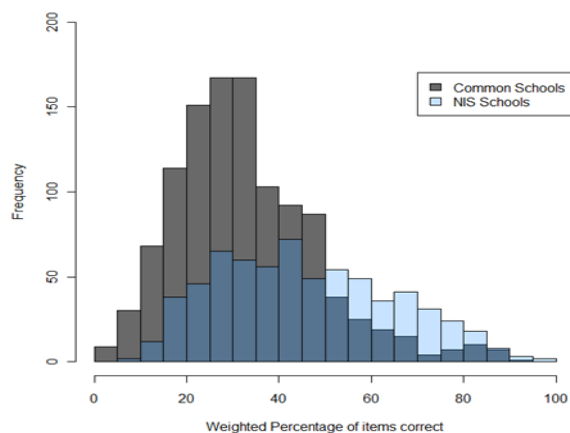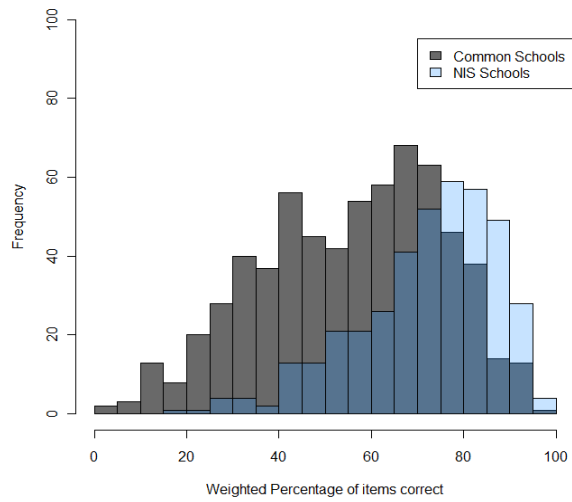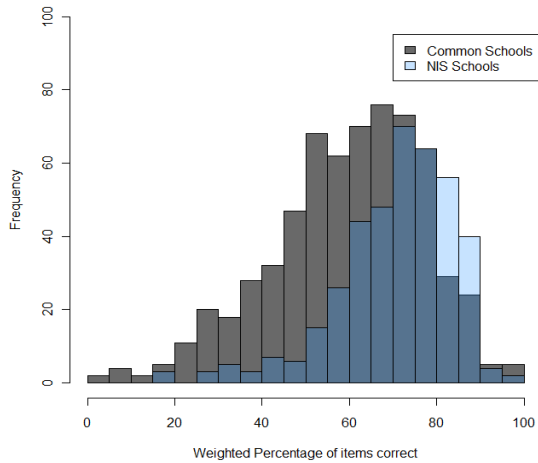

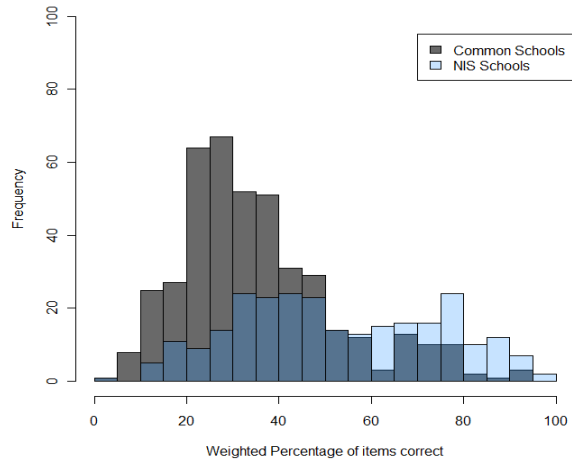Mathematics   English

Russian as first

Russian as second

Kazakh as first

Kazakh as second

SELECTION RULE

The performance standard set by experts turned out to be very high and unrealistic. Pending further research in setting a proper content performance standard, it was decided to set the minimum so high that the probability that a candidate would achieve this score by pure guessing would be minimal. To calculate this, the test characteristics from the pretest have been used. The minimum score set for Mathematics is (35%). The results of students that qualify according to the minimum score on Mathematics are ranked according to the total scores for all four subjects of the test.

The test should offer a fair opportunity to all test-takers. The content of the selection test is a careful balance between mathematics, foreign languages and mother tongue and is meant to trigger the student to show his or her capabilities and abilities. Through the composition of the selection test, emphasis is put on mathematics by including 40 items to 20 items for the language subjects. The total score of the subject test, therefore, places more emphasis on mathematics. The total score was found to be suitable for ranking the students.

Please not that both the NIS 'subject' test as well as the CTY test are used in the selection of the candidates.

JUNE 2013 SELECTION TEST

In June 2013 one of the pretested versions was used in the selection test. There were 5277 candidates in total participating of which 4050 candidates (77%) and 1227 candidates (23%) with Kazakh and Russian as languages of testing respectively.

Test results analysis has showed that the minimum requirement set for Mathematics Selection test has coped well with the task of discriminating the most able students from less able students. In average 42 % of Kazakh stream candidates and 60% Russian stream candidates qualified as eligible according to Math cut off score.

Main findings on test characteristics

In table 2 and 3 below the main test characteristics are presented. The characteristics are given for the total test as well as per subject for the Russian version in table 1 and for the Kazakh version in table 2.

The coefficient Apha for the total test was 0.85 for the version for Russian as language of instruction and and 0.82 for the version with Kazakh as language of instruction. These values comply with criteria for high stakes test. But we would like these values to be higher. There are possiblities for improvement: a few items did not function well; they were not discriminating well. They cousd be replaced by ones with a better discrimination. The average p-value for mathematics is a bit low and the average for Kazakh and Russian a bit too high. To make the test more discriminating the average p-value should be lower. In selecting items for future versions of the selection test NIS will take these finding into consideration.

Also new items of a more appropriate difficulty level will be constructed in the near future.

Table 2 Test characteristics Russian

	total test	Subtest(s) math	kaz	rus	eng
Number of testees	1230	1230	1230	1230	1230
Number of items:	100	40	20	20	20
Average test score:	**52,92**	**16,02**	**14,19**	**13,05**	**9,66**
Standard deviation	11,56	5,93	2,81	4,43	3,84
SEM:	4,46	2,83	1,81	1,83	2,03
Average P-value:	**52,92**	**40,04**	**70,96**	**65,24**	**48,32**
Coefficient Alpha:	0,85	0,77	0,59	0,83	0,72

Table 3 Test characteristics Kazakh

	total	Subtest(s)			
	test	math	kaz	rus	eng
Number of testees	4057	4057	4057	4057	4057
Number of items:	100	40	20	20	20
Average test score:	**47,3**	**14,26**	**14,17**	**11,19**	**7,68**
Standard deviation	10,5	5,06	2,95	3,74	3,19
SEM:	4,49	2,8	1,92	1,95	2,05
Average P-value:	**47,3**	**35,64**	**70,86**	**55,97**	**38,38**
Coefficient Alpha:	0,82	0,69	0,58	0,73	0,59

FURTHER RESEARCH

For further development and improvement of the system, it is planned to carry on research in the field of investigating the most appropriate and adequate minimal requirements (cut off score), which enable to select students, that will have academic success in Intellectual schools and will build nation's intellectual capacity. In addition, we are aware of the need to develop items of appropriate difficulty level, that assess more higher order skills. This is the challenge that we are ready to face, should go in line with the development of the curriculum that delivers these higher order skills to students.

List of Reference

Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.
Downing, S.M. &Haladyna, T.M. eds. (2006). Handbook of test development. New Jersey: Lawrence Erlbaum Associates, Inc.
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/