Progress in the validation of high stakes assessment: the case of Selection test for Nazarbayev Intellectual Schools in Kazakhstan

Rustam Abilov, Miras Baimyrza, Dariya Kulbayeva, Zamira Rakhymbayeva (NIS, Kazakhstan)

Nico Dieteren, Frans Kamphuis, Frans Kleintjes (Cito, The Netherlands)

Abstract

INTRODUCTION

Nowadays, teaching and learning in many countries are oriented in acquiring competences by students to enable them to function as citizens in modern society. This orientation is often reflected in the concept 'Functional literacy' which is a quite popular terminology used by many researchers. The concept that was almost for the first time introduced by the developers of the PISA project. Worldwide, the modern school should prepare students (and we are intentionally trying to skip the word 'teach') to not only know things but also to be able to apply that knowledge in real-life situations.

Also in Kazakhstan there is a need to prepare students for future life, not only from the perspective of the student, but also from the perspective of the nation. Nazarbayev Intellectual schools (hereinafter referred to as NIS)) were initiated with the aim to build an intellectual capacity of the nation, focusing on the representatives of the new generation who will contribute to the prosperity of the country in the near future. To establish this goal a new curriculum has been developed. A curriculum that meets the demands of the society. The implementation of the new curriculum started in September 2012.

Implementing a new curriculum asks for an appropriate selection procedure to be sure that NIS students selected will be able to master the content of the new curriculum.

The previous selection procedure resembled a classical post-soviet approach in selecting students: a combination of elements that remained from socialism and elements that were gained during the new era. In the past system a two-stage test procedure was used to select potential

students. Students who pass the first stage were invited to take part in stage two. Based on the results on stage two it was decided whether a student is allowed to study at NIS. The first stage consisted of items related to Mathematics, Kazakh language, Russian language and English. Depending on the stream (Physics and Mathematics or Chemistry and Biology) a student would like to choose, mathematics could be replaced by chemistry, biology or physics. Each subject is represented by 15 items in this first stage. If the applicant achieves a passing grade on the first stage, he or she can participate in the second stage, which consists of 5 complex mathematics assignments, including higher order skills assignments. A minimum score on this second stage is needed to pass. If a student's result is close to the cut-off score, additional information about the student is used for the final decision.

There were three general problems with the past test. *Firstly*, the test is very much focused on knowledge. This means that applicants can study for the test and that the test is necessarily curriculum-dependent. For a selection test it is important to select applicants that have the necessary general ability or intelligence to succeed in NIS. A higher skills test, which focuses less on knowledge itself and more on the application of that knowledge, would be more suitable for this purpose. *Secondly*, the open questions in the second stage require markers. As it is a high-stakes test this can lead to security risks - more people see the questions - and have a potentially lower objectivity and, therefore, also a lower defensibility. *Thirdly*, there is no rationale for the cut-off score. An arbitrary cut-off score threatens the validity of the scores and also the defensibility of the test. In view of these problems it was decided to replace the current selection test by a new test.

The selection procedure needed revision in order to take new developments into consideration. Starting from 2011, given the new curriculum, a revised selection procedure was developed to aim at assessing abilities and application of knowledge in order to select the most potential students, rather than assessing knowledge.

Current Selection procedure

The students of NIS are now selected both upon the results of academic achievement and ability. The selection test is standardized test, administered in two days. First day for a subject test on

academic achievement in mathematics, Kazakh language, Russian language and English language was developed conjointly with *Cito Institute for Educational Measurement (the Netherlands)[1]*. Second day for an ability test that contains a quantitative section and a section that assesses spatial reasoning produced by *The Johns Hopkins Center for Talented Youth (USA)*.

For the selection procedure it means that it has changed in such a way that it will select those students that will be able to acquire the skills and knowledge as indicated by the curriculum and predicting successful learning in NIS schools. This means a thorough change in the type of questions that are asked in the selection test: more application of knowledge and more higher order skills than (just) knowledge. The selection test should aim at assessing potential rather than remembering or knowledge.

Using sophisticated methods for psychometric analysis of test data, including IRT (One Parameter Logistic Model) enables NIS to offer a fair, objective, transparent and fully justifiable procedure for selection, throughout all the regions of the country and on several testing moments and locations.

Students have to get minimum score on Math (140 or 35% out of 400) and Quantitative (24 or 40% out of 60) tests to be ranked for further selection. Limited number of scholarships per school asks for ranking student test results for a fair scholarships award. Firstly, candidates are ranked according to their total score on subject test. If total score on subject is equal, students are ranked by their score on Mathematics. If score in Mathematics is also equal, students are ranked according to their total score on ability test.

Total score on subject test		
	Math score	
		Total score on ability test

---

[1] please see "*Subject Knowledge Assessment in the Selection of Students to grade 7 of Nazarbayev Intellectual Schools*" presented at 39th IAEA Annual Conference Tel-Aviv, Israel, October 20-25, 2013 for more information

In 2013 the first selection test was administered at 9 schools in different regions of Kazakhstan. In 2014 selection test was administered 4 times in 20 schools with 11 828 candidates. 18% or 2 236 candidates out of 11 828 candidates were awarded a scholarship for education at 7th grade at NIS starting from 2014-2015. Data from this cohort will be used for analysis.

Table 1.

	2013	2014	2015
Number of schools	9	20	20
Number of candidates	7454	11 828	13540
Number of scholarships	1000	2236	2415

Purpose of the Study

For this very high stake assessment process all is done to proof to all stakeholders in Kazakh society that the children with best intellectual potentials are the ones selected. In accordance with the United Nations Convention on the Rights of the Child, each child has to be treated fairly, bearing in mind the best interests of the child. Therefore, test administrators must see the best interests of children as a primary concern in any decision making process that may affect them. In doing so, test owners have to do their best to present transparently how test is being used and administered. This view aligns with the Standards for Educational and Psychological Testing (AERA/APA/NCME, 1999), which states that the test maker is "responsible for furnishing relevant evidence and a rationale in support of the intended test use" (p. 11). Perhaps one of the first steps in test evaluation is validation process.

Validity evidence comes in many forms, where paper intends to build a study based on Cranbach's (1971, p. 443) definition, where "validation is the process of examining the accuracy of a specific prediction or inference made from a test score". Therefore, we consider validity argument as main evidence of test implementing the purpose of the test. Purpose of the test is selection of most able students for successful study at NIS, apparently one of the critical forms of validity is evidence of test's predictive validity. As a result, the purpose of the present study is to determine to what extent selection test predicts future success at school.

METHOD

Population and Data collection

The population of interest consisted of grade 7 NIS students, who were selected in 2014. In 2014, 11 828 students participated in both days of the selection tests. Approximately 1965 out of 11 828 students were awarded scholarships to 17 schools throughout the country. They have started grade 7 in September. As part of their education at NIS, these students participated in two moments of monitoring math (September and January), as well as in monitoring second and foreign languages (April).

Monitoring math in September consisted of five domains, such as Algebra, Numbers, Geometry, Statistics and Modeling; in January there were three domains, such as Numbers, Geometry and Modeling.

Monitoring of second and foreign languages is administered for Kazakh as a second language, Russian as a second language and English language. Those students, who are educated in Kazakh language stream did test in Russian as second, whereas students in Russian stream di test in Kazakh as a second. At the same time English language is considered as a foreign language for both of streams, and as a result was administered for both of the language streams.

Initially data was collected for 1 814 students from different databases maintained by different project teams.

Variables and Indicators

Variables: selection test results

Criterion variable: results of monitoring math, monitoring languages, school based assessment

Data Analysis

Research questions one

To what extent do Selection test scores taken singly (individual Mathematics and Quantitative), and total score on Subject test, predict success at school as measured by Monitoring Math results

Research questions two

To what extent do Selection test scores taken Kazakh as second language, Russian as a second language, English language, predict success at school as measured by monitoring test of reading skills in Kazakh as a second, Russian as a second and English languages.

Research question three

To what extent do Selection test scores taken singly (individual Mathematics, Quantitative, languages, Spatial ability), and separately total score on Subject test and total score on Ability test, predict success at school as measured by grades for Kazakh language, Russian language, English language, Mathematics, Chemistry, Physics, Biology.

RESULTS AND DISCUSSIONS

CONCLUSIONS AND RECOMMENDATIONS