Validity of the assessment approach in the monitoring system for languages

Rustam Abilov, Assem Issabekova, Anara Dyussenova (NIS, Kazakhstan), Jesse Coops (Cito, The Netherlands)

INTRODUCTION TO THE MONITORING SYSTEM IN NAZARBAYEV INTELLECTUAL SCHOOLS

Education in Kazakhstan is facing a challenge of implementing trilingual policy at school level so that three languages (Kazakh, , Russian, English) will be at fluent use in the future. Nazarbayev Intellectual Schools have been developing various opportunities for students, teachers and parents to reach this aim since 2008. An ambitious yet challenging goal, initiated by President of the Republic of Kazakhstan, gave life to Nazarbayev Intellectual schools, abbreviated as NIS (Autonomous Educational Organization Nazarbayev Intellectual schools' 2020 development strategy, April 2013). In general, this new stream of NIS schools has brought the development of new curriculum, analysis, research and monitoring tools that include best international and national practice in education.

The trilingual policy is based upon the belief that students can develop not only their knowledge at school but also become intellectually developed individuals capable of using various practical skills learned at NIS in their real life. In NIS this is achieved through language competency not only in their first but second and even third languages, available to students through content based subjects. Team teaching is a new way in language instruction to help students to acquire several languages at a time. In addition, it is a combination of various resources such as integrated textbools, curriculum, sofware, technical equipment, trilingual library collections as well as opportunities to learn and practice languages with native teachers. This is a long and not full list of possibilities that assist and guide learners in achieving the ambitious goal.

To inform NIS administration, parents, students and teachers if they are developing the capacities to meet the goal required a diagnostic tool that will take a moment in the learning process like a shot. This led to the development of the monitoring system for languages in NIS that started in 2013.

NIS has been cooperating with one of the leading organizations in the field of assessment, Cito Institute for educational measurements (the Netherlands). A longterm project was set to develop a monitoring tool that will help to obtain reliable information about students' immediate perfomance in the four known language skills, namely listening, reading, writing and speaking. The information is very important to diagnose "where we are now". NIS and Cito have been working in this direction for more than 2 years. Teachers from NIS schools is another important component of this cooperation.
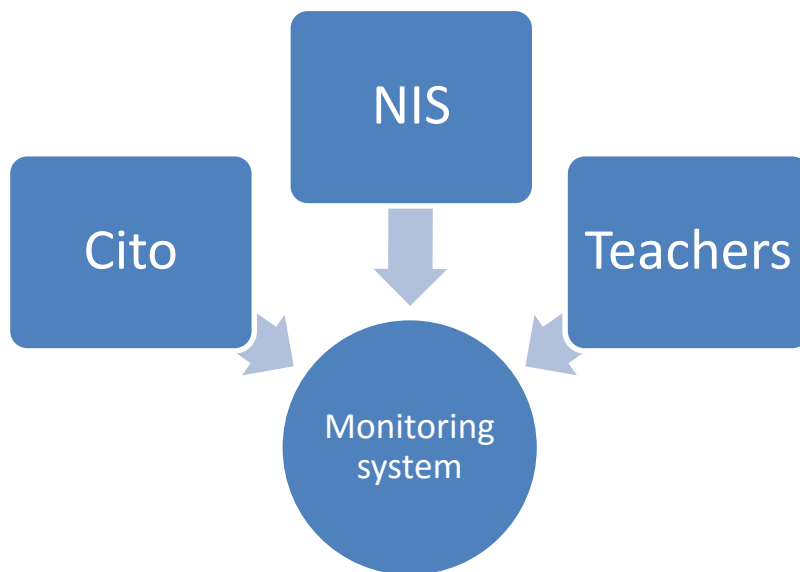
During this period, many teachers from NIS schools have grown to item-writers, developing and improving their skills on various training workshops provided by Cito. They have learned to develop test items and tasks, which are always pre tested before being implemented to a real monitoring test.

THE MONITORING TEST

Reading and listening skills contain multiple-choice questions, while writing and speaking are tested individually. Compared to straightforward administration of the listening and reading tests, speaking and writing are quite time-consuming. Most importantly is that teachers assess writing and speaking skills. They evaluate students' ability in speaking via face-to-face interview; two teachers evaluate while one teacher interviews the students. Teachers also assess students' written works. This scheme has been developed to ensure the validity of the assessment system. The principle is that teachers' marks should be more or less common. The assessment procedure is strictly followed by a specifically developed online marking scheme for the two skills.
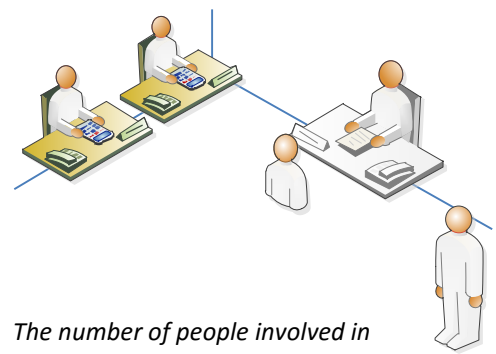
Responses to speaking and writing tasks have to be marked on several criteria. For each skill, there are two types of marking scheme: one type for the lower level tasks (A1-A2) and a different type for the higher level tasks (B1 and higher).

SPEAKING

The speaking test is probably the least reliable test in the monitoring system. Reliability can be improved by training of raters and the use of smaller subtasks. In this way an acceptable reliability can usually be achieved.

A second factor which influences students' test scores is the severity of raters. Even if inter rater reliability is high, there can still be relatively large differences in scores awarded by different raters for the same students. According to Bejar (1985) the severity of raters often differs widely. Awarding the mean score of multiple raters is suggested as a compensatory device.

To administer a speaking test an interlocutor and at least one rater are required. As was mentioned above, it would be most prudent to include a second rater to improve reliability. It is also advisable to have a fourth person present, to guide the students in and out, and have a rotation schedule so the fourth person can relieve one of the raters or the interlocutor to prevent rater-fatigue. An alternative for the second rater or even both raters would be to record the speaking test to be scored later. This

The number of people involved in speaking assessment of a single student

would need to be a video-recording since gestures and expression are a part of the CEFR.

WRITING

As with the speaking, writing is also assessed by using a marking scheme completed by two raters. To make it more reliable students' written works receive two independent marks.

TEST ITEMS AND TASKS

For speaking and writing open (constructed response) tasks are being developed. They are constructed, reviewed, adapted and pretested each time before becoming part of a real test. Pretest results usually show how good and qualitative these tasks are. The administration and marking of these tasks is not as straightforward as for listening and reading. For this reason, a larger group of language teachers are trained in administering and marking these tasks. These teachers then further instruct other language teachers in their own schools.

THE MONITORING SYSTEM

The monitoring system developed at NIS is not straightforward and as a whole it is being tested each monitoring moment, but it clearly gives interesting and useful information needed to make further decisions on students' development in their first, second and third languages. Validity of the approach to testing the four skills, especially speaking and writing is of main concern however, it is achieved through trainings, technologic resources, adequate test construction and accurate marking schemes.