Ways of efficient use of data produced by integrated assessment system for individual, institutional and nationwide decision-making

Authors:

Maleyka Abbaszade, Chair of the Board of Directors, State Examination Center of the Azerbaijan Republic; E-mail: m.abbaszade@dim.gov.az

Parviz Kazimov, Head of Strategic Development and Project Management Unit, State Examination Center of the Azerbaijan Republic; E-mail: p.kazimov@dim.gov.az

Keywords: assessment, examination, decision, data

Abstract

The State Examination Center (SEC) of the Azerbaijan Republic manages complex assessment system including secondary and high school final examinations, university entrance, as well as selection of candidates for civil service.

While in early stages of its operations the only purpose of the centralized examination system was ensuring impartial selection of students for universities, later it enhanced its technical and research capacity transforming into huge data production platform for individual and institutional decision-making. The platform has been designed to align requirements of all stakeholders including candidates and institutions.

Flexibility, technology-based design and scientific approach enables the system to act as a facilitator and from time to time even a driving force of positive changes and innovations aimed at individual and institutional capacity building.

Network of experts and educators managed by the SEC develops various assessment tools, administers assessment procedures in line with specific requirements of education institutions, government agencies and individuals. Education institutions and government agencies can use the produced data to revise their strategies and develop human resources, while individuals can take advantage of it for education and career planning.

In 2016 SEC started to administer the selection of candidates to civil service, which opens a new perspective for validating assessment on higher education level and reconciling the university education with real career requirements.

This report provides the logical analysis of the data produced by different branches of integrated assessment system and shares experience of using this data for individual, institutional and nationwide decision-making. It also informs on methods and tools used for data collection and ideas for further improvement of the system.

Introduction

After the collapse of the totalitarian society in last decade of 20th century Azerbaijan, like most former soviet republics, faced the challenge of transition from system-centered to individual-centered education system. Unlike the soviet system, the main purpose of which was cultivating citizens in line with official ideology, the education strategy of the newly independent state

intended to provide more freedom for individuals in terms of personal decisions and career choices. As later developments showed, the centralized university admission system meant to play one of key roles in this transformation.

Introduction of centralized exam system in early 90s was welcomed by the society as it was considered an effective antidote against corruption, another by-effect and remnant of the totalitarian system. The questions in the heads of people regarding transparency of the process and impartiality of the results were quickly solved by professional approach of the staff and especially application of information technologies. Maximum automation of the process, including development of test pool, calculation of scores and assignments based on clear-cut standards facilitated the management, ensured transparency and consequently gained public trust in early stages of life of the system. In later stages, however, when the problem of corruption seemed to be left behind, the questions rose regarding effects of the system on school education. High competition for popular university programs and insufficient quality of education in general education schools gave rise to so called new sub-industry in education sector – the tutoring. Numbers of experts and officials complained that the teachers and tutors tend to adjust the teaching process to the model of high-stakes exams which was solely based on multiple-choice questions. Thus the problem appeared much more complex than it seemed in first years of application of centralized exams, which includes social, cultural and economic aspects. Regular attempts of corruption including mass cheating and illegal admission of students that were traced and exposed thanks to the sophisticated algorithms applied in admission system and regular control measures showed perils of loosening the decision-making process or returning to university-based selection. Nevertheless the new situation created new challenges which must have been addressed. The SSAC (predecessor of SEC) in cooperation with the Ministry of Education choose to go forward by diversifying tools and content of exams and providing more options for students.

Introduction of centralized exam model for university entrance

Application of centralized assessment model started with the introduction of test model for admission of students to universities. All university programs were placed in five (eight groups in earlier stages) specialty groups and the selection of students were administered through test exams with unified content for each group. Applicants who scored higher than criterion-referenced minimum cut-off score aimed at formation of competition pool could select 15 programs (3 programs in early years) included in the same group and final decision was made based on second cut-off score (norm-referenced) determined by the computer algorithm based on applicants' scores and number of openings for each program. If the applicant's score was not sufficient to enter to the program of his/her first choice, the algorithm automatically moved him/her to next choices until the score allowed winning opening in the particular program as a result of current competition conditions. Applicants to programs that required specific skills beyond general knowledge like sport and arts should also pass specific tests administered by the university itself under the scrutiny of SSAC (SEC).

In year 2008 the Ministry of Education initiated transition to curriculum system in general education. In such a case the university entrance exams should have been adapted to the curriculum to ensure valid assessment. The SSAC started development of new assessment tools, including constructed response questions, open-ended questions and essays for assessment of

students trained based on curriculum model. As the curriculum-based training in general education schools was provided only for the students who started the school after year **** and the older children continued to study with traditional program, transition to the new exam model was administered in several phases. Full-scale application of this model started in 2019 along with substantial changes in admission model. But earlier the limited number of open-ended questions was included in the university entrance test to push teachers to focus on stimulating problem solving skills of students and to prevent misuse of multiple-choice questions during lessons and tutoring sessions. The analysis of the test results reveal that open-ended questions make positive contribution to discrimination among high scorers.

Final exams in general education

In year 2012 the government decided to introduce centralized assessment of students in general secondary (9^{th} grade) and complete secondary (11^{th} grade) education stages. The tests for 9^{th} and 11^{th} grades included both open-ended and multiple-choice questions on two (Math and Native Language) and three subjects (Math, Native Language and Foreign Language) respectively. Initially the results of final exams were not used for any high-stakes decisions for students; instead they were regarded as monitoring tools to collect detailed data on the quality of education in general education schools and for transition to the new model. From the year 2015 the results of these exams for the first time were used for selection of students to secondary special schools and for assessment of general skills of candidates to university programs included in V group (arts, sport etc.) in addition to the assessment of specific skills (aptitude test).

Implications for education authorities

Comparative analysis of the results of Grade 9 graduation exams reveal that the students' perform better with multiple-choice questions than with open-ended. Even though the students that took the Grade 9 graduation exam have been studying based on the curriculum that develops their skills and abilities from Grade 1, we see that they faced difficulties in solving these tasks.

The score (in percentage) of an average student in open-ended test questions was 30.05 and 55.98 in closed-ended test questions. These analyses lead us to conclude that implementation of curriculum in general secondary schools is quite poor, and the skills and abilities promoted among children is not at the appropriate level.

Correlation coefficients by subjects between 1st (mostly open-ended questions) and 2nd stages (multiple-choice close-ended questions) of the Grade 9 graduation exam were also calculated. There is a statistical relationship, which can be considered significant. The general correlation coefficient between the results of the 1st and 2nd stages by the language of instruction subject is equal to 0.79. The correlation coefficient is equal to 0.73 for math.

Correlation coefficients between 1st and 2nd stage results (Grade 9)

Subjects	
Language of instruction	0.79
Math	0.73
Total including both subjects	0.86

This means that difficulty in answering open-ended questions that require problem solving skills rather than memorization is true for the whole group of examinees irrespective of the performance level.

The comparison of distribution of results obtained by Grade 9 and Grade 11 students in graduation exam subjects shows that the results of Grade 11 students are better. An average Grade 9 student scored 51.73 points in language of instruction subject, 47.48 points in math subject in the 2nd stage of the graduation exam. An average Grade 11 student scored 57.95 and 55.95 points, respectively. Hence, Grade 11 students have performed better compared to Grade 9 students. It leads us to conclude that the students take education more seriously after Grade 9. We think that the main reasons for this are the students' attending ***extracurricular*** activities in order to prepare for entrance exams to higher education and specialized secondary education institutions on the basis of complete secondary education, and their additional preparation with tutors. The finding also implies that the students are motivated more extrinsically than intrinsically.

Current university admission model

Within the current model for admission of students to universities launched in 2019/20 academic year the Grade 11 final exams are considered as first phase of two-phase university entrance exams. Graduates of earlier years have also to take the similar exam. As mentioned above the test includes 85 questions in total including 30 on native language (Azerbaijani or Russian), 30 on foreign language and 25 on math. The maximum possible score is 300 points – 100 for each subject.

The result of the exam is valid for next two years and allows the applicants to participate in all eligible competitions during current and next academic years. Those who are discontent with their result can retake this phase next year.

The second phase of exams which is called university entrance examinations are based on specific content for each specialty group. The results of the entrance exams are valid for only current year and appropriate specialty group. Unlike general education final exam the university entrance exam is held twice in a year (spring and summer sessions). The applicant can take test for different specialty groups in each session if he/she desires. As the first phase is universal, taking admission exams for two groups the applicant can compete for university programs included in two different groups. This allows the applicant more options in terms of program selection. The applicants can also chose to go for V group programs by passing specific aptitude test administered by the university. The applicants who scored minimum cut-off score in first phase (or Grade 11 final exam) are permitted to take part in aptitude test for V group university programs.

Entrance exam test for each of I-IV specialty groups includes total 90 questions split into three batteries for relevant subjects. Each battery includes 30 test questions. Correct answers for each subject are assigned 1 or 1.5 weight coefficient depending on considered importance of the subject for the programs included in the group. The total possible score for entrance exams is 400. So the applicants can score 100 and 150 points for the subjects with 1 or 1.5 coefficient respectively.

The contents of university entrance exams for each group and weight coefficient for correct answers have been determined as below:

Specialty groups		Math	Physics	Chemistry	Geography	History	Native language	Biology	Aptitude exam
I group	Math, physics, engineering, architecture	1.5	1.5	1					
II group	Economics, management, international relations, regional studies, sociology and geography	1.5			1.5	1			
III group	Humanities, pedagogic specialties	1				1.5	1.5		
IV group	Medicine, chemistry, biology, psychology, agriculture		1	1.5				1.5	
V group	Fine arts, music, sports								*

Final exam as multi-purpose tool for decision-making

As we see the school final exams have been given a universal role for selection of students to different types and levels of higher and special education. Students can apply for secondary special schools only with scores of school final exams. Many programs for secondary special education are available for students after Grade 9. According to the latest amendments to the education legislation the graduates of secondary special schools (after 9 or 11 grades) are granted the sub-bachelor degree and given a chance to participate in competition for undergraduate programs. Sub-bachelors are automatically considered passing first cut-off score in order to be included in competition pool. Competition algorithm gives preference to the applicants who took entrance exams in current year over sub-bachelors. Sub-bachelors can compete only for vacant seats of university programs that were not applied for by applicants who passed entrance exams in current year.

Why not certificate grades, but graduation exam is used for decision-making?

Looking from outside a question can rise regarding current university admission model: why not school grades but the final exam results are considered for decision making?

The average certificate grades of applicants based on marks are indicated in the application form and this enables to make comparison with their scores in entrance exams. The SEC conducts comprehensive analysis of results after each years' university admission campaign. Comparative analyses reveal that grading in most general schools of the country cannot be taken as valid data for decision-making. For instance, in 2018 only 52 percent of applicants who finished schools with "5" (highest) grades were able to justify their certificate grades. In such a case, if any additional point is added to the applicant's result during the admission to higher educational

institutions based on considering the applicants' certificate grades, so called "high achievers" that don't justify their excellent marks can take unfair advantage over those who score points based on their knowledge in admission exams. On the other hand, it is difficult to guarantee that such a step wouldn't lead to more subjectivity. That's why we do not support the idea of considering certificate grades as an additional data for decisions until there is a high correlation between school grading and admission exam results.

On the other hand comparison of results of Grade 11 final exams and university entrance test reveal quite positive correlation. The nature of stochastic relations between the results of the same applicant both in graduation and entrance exams based on graduation and entrance exam results was studied. The correlation coefficients calculated based on ability indicators defined according to the number of correct answers and the two parameter logistic model (BILOG-MG software) are presented in the table below.

CORRELATION COEFFICIENTS BETWEEN GRADUATION AND ENTRANCE EXAM RESULTS

SPECIALITY GROUP	Mother Tongue	Math	Two subjects together
I	0.802 (0.788)*	0.735 (0.738)	0.832 (0.827)
II	0.813 (0.807)	0.734 (0.745)	0.830 (0.839)
III	0.841 (0.813)	0.758 (0.779)	0.850 (0.857)
IV	0.828 (0.791)	0.828 (0.771)	0.875 (0.845)
TOTAL	**0.819 (0.804)**	**0.745 (0.755)**	**0.838 (0.840)**

** The correlation coefficients calculated based on skill indicators defined according to two-parameter logistic model of the modern testing theory are provided in parentheses.*

As seen from the table, there is a significant statistical relationship between graduation and entrance exam results. Such a question emerges when we take these indicators of the correlation coefficient: can the results of only one test be used both for final attestation and student admission to higher educational institutions? Comparing information functions of both tests, we see that these tests were intended for addressing different issues. Hence, entrance exam tests do not ensure high accuracy in assessment at the low skill interval. As for graduation exam, it is not sufficiently effective in assessing at the mid- and high-level skill interval.

So the Grade 11 final exam results are good for selection of students for secondary special schools and add value for selection of students for relatively low rated university programs, while data produced in university entrance test useful for selection of students to highly popular university programs.

Options provided to applicants by current university admission model

Current university admission model provides several routs to college and undergraduate programs. Applicants can enter secondary special education schools with the results of either Grade 9 or Grade 11 final exams. Grade 11 exams are considered also a first phase of university entrance examinations and the result is valid for current and next year admission campaign. University entrance test is administered twice a year providing the applicants a second chance to improve their result in case if they are not happy with the score gained in first test or to apply for another group of specialties if they wish. The applicants can also take aptitude test to enter

university programs requiring specific abilities included in V group of specialties. The graduates of secondary special schools also have chance to compete for some undergraduate programs without any exam. Application to the programs start after announcement of results. The SEC also publishes some data from previous years admission campaign like cut-off scores for programs, admission plans and number of applicants in particular score range. So the applicants can make decisions with variety of data under hand that enable to estimate their chances for particular program.

Advantages of current model for higher education institutions

The tests used in general school final exams and university entrance exams include portions of constructed response and open-ended questions which assess problem-solving, analythical and other practical skills. This complements curriculum approach and stimulates students and teachers to focus on development of respective skills rather memorizing the material.

The fact that unlike earlier years this year no applicants scored maximum 700 points in aggregate indicates that the new model puts new challanges for the students. This will push the students and teachers to revise their approcah to preparation for entrance tests and as a result in coming years we may have more skillful students in universities with better cognitive skills and internal motivation.

Observations in universities revealed that students of programs instructed in english have some difficulties with language profficiency especially in terms of listening skills which hinders the efficiency of the study. Starting from this academic year the SEC introduced listening tests in entrance examinations for both undergraduate and graduate programs. We beleive that this will stimulate teachers and students to take more practical approach in foreign language lessons.

Data produced for national policymakers

The SEC annually developes data on the basis of results of examinations that can be used for education policy planning. Annually published report includes comprehensive data on various aspects of the education, including results of education departments (regions), performance of separately taken schools, gender analysis. For instance, comparative gender analyisis of the results of Grade 9 and Grade 11 exams reveal that proportion of girls seriously drop after Grade 9 exams in some regions. It may imply some socio-economic factors, as well as a problem like early marriges. Relatively poor performance of students on test items that require problem solving, data analysis draws attention of the education authorithies to qualification of instructors and efficiency of methods used in general education.

Centralized selection of students for graduate programs

In year 2005 the centralized exam model started to be applied for admission of students to graduate programs including medical programs. The model included two-phase assessment. The first phase of the exam tests knowledge level in logical thinking, foreign language and informatics through multiple choice questions. Those who meet the defined competition requirements are qualified to participate in the second stage.

In the second stage, the knowledge of undergraduates is tested through 50 test questions presented to them. Test for each program includes 5 open-ended questions. Undergraduates that selected specializations under certain programs have a written exam instead of test questions. The undergraduates who want to be admitted to master's programs that require special aptitude took an ability test instead of answering test questions for their specialty in the second phase of admission exam.

Competition rules for admission to graduate level specialties

Name of specialties groups (specialty) in Master's level	Minimal requirements for results at the first stage of exam*			Minimal requirements for results at the second stage of exam		Minimal requirements for general score	
	Minimal requirement for subjects		Minimal requirement for stage	For State Order seats	For paid seats	For State Order seats	For paid seats
	Subject name	Minimal score					
Group of educational specialties Group of technical and technological specialties Group of health, welfare and service specialties (Except for physical education and sports specialties)	Development level of general intellect (logical thinking)	10-15 (depends on the program)	**30-50 points (depends on the program)**	**20-25 points (depends on the progam)**	**15-20 points (depends on the proogram)**	**50-75 points (depends on the program)**	**45-70 points (depends on the program)**
	Informatics	5					
	Foreign language	5					

Admission exams in Residency (medical programs) are also conducted in two stages. The first stage includes an exam that assesses knowledge on core subjects and the second stage exam focuses on assessment of knowledge on major subjects. Those who scored at least 40 points in the 1st round of the exam qualified to the 2nd round. The candidates were presented 100 test questions separately in each stage. The tests include both multiple-choice and open ended questions. Each correct answer to test question was marked as one point. Wrong answers didn't affect the candidate's exam results.

The data collected in two critical stages of the education process allows the SEC (SSAC previously) to develop comprehensive reports reflecting quality of education both in general and higher education institutions of the country.

Statistic relationship between diploma points and master's admission test scores

Correlation indicator that reflects a statistical relationship between average diploma points of undergraduates that represent each higher educational institution and the points scored in the first stage of master's admission exam was calculated. Analysis of correlation coefficients calculated for each higher educational institutions shows that, taking as a whole, the correlation between score indicators is weak (0.43). The value of correlation coefficient among respective score indicators of undergraduates representing majority of higher educational institutions ranges between 0.33 and 0.58.

In 2018, 14 test tasks on verbal reasoning, 15 on spatial reasoning and 21 on logic and math were presented to undergraduates in order to test the development level of general intellect (logical thinking) at the first stage of Master's admission exam.

Statistical analysis of undergraduate exam results on test question that identify development level of ***logical reasoning*** shows that more than 70% of the undergraduates have failed to answer 17 out of 50 questions. As for 26 test questions, this indicator ranged between 20-70%. The indicator on correctly answering 7 questions was below 20%.

	Number of presented test questions	**Indicator on correct answering**		
		Less than 20%	**20 - 70%**	**More than 70%**
Verbal (of words) test questions	14	4 (including 1 open-ended)	7	3
Test questions on shapes	15	0	6	9
Logical and mathematical test questions	21	3 (including 2 open-ended)	13	5
TOTAL	**50**	**7**	**26**	**17**

A small portion of the undergraduates that took the test was able to answer open-ended questions correctly (2018). Analysis shows that while the indicator to correctly answer closed-ended test questions that test development level of logical reasoning was on average 57.41% this year, the same indicator for open-ended questions was only 9.77%.

The foreign language test unit mostly included questions that were communication-oriented, required logical approach, skills to make generalizations. The tasks allowed to reveal the knowledge, skills and abilities gained by undergraduates in foreign language learning.

Scientific and methodological analysis of the exam results by all foreign language subjects and sectors (Azerbaijani and Russian sectors), as well as difficulty level and ability indicators of test questions allow us to say that most of the undergraduates have a command of foreign language that is close to intermediate. However, these results differ according to different languages and subject topics.

Analysis of used test questions on their difficulty levels shows that most of the undergraduates performed well in answering easy and mid-level difficult fact-based test questions. They performed relatively poorly in answering medium-level and difficult test questions that were application-oriented, required critical thinking, generalization skills.

Undergraduates' percentages of answering closed-ended questions correctly (average indicators)

	English	**German**	**French**	**Russian**
Azerbaijani sector	50.38	52.84	53.99	55.98
Russian sector	58.13	57.89	61.50	–
Total by subject	51.27	53.41	54.19	55.98

The undergraduates' answers to open-ended questions in the foreign language test section were not that satisfactory. The results are presented in the table below.

Undergraduates' percentages of answering open-ended questions correctly (average indicators)

	English	German	French	Russian
Azerbaijani sector	30.62	26.49	23.65	32.60
Russian sector	35.08	27.86	43.33	–
Total by subject	31.13	26.64	24.18	32.60

Conclusion

Integrated assessment system in Azerbaijan produces data for decision-making in individual, institutional and national levels. Comprehensive analysis of results of annually administered examinations shed light on to the problematic areas in the education sector. Continuous development and diversification of assessment tools and methods contribute to better selection of students, as well as stimulation of individuals for self-development in line with contemporary demands of the labor market.

Azerbaijan moves forward by strengthening and continuously improving its centralized examination system that has been proven efficient by decades of experience. The main challenge for SEC in specific social and cultural conditions is that it should develop proper capacity to provide more options and freedom for individuals in decision-making at the same time to preserve general control over quality of the assessment.

Careful planning of next steps based on scientific approach and regular data analysis in cooperation with education authorities and institutions are considered as important elements of innovations. Thousands of experts and instructors annually participate in workshops organized by the SEC to make contribution to the improvement of the system. We also value international expertise in this field and happy to participate in international events like IAEA annual conferences to take advantage of international experience as well as to share our experience with the partners.