The Development of the Student Assessment System in the Republic of Armenia: Achievements, Challenges, and Lessons Learned

George Bethell and Karine Harutyunyan
The Development of the Student Assessment System in the Republic of Armenia: Achievements, Challenges, and Lessons Learned

George Bethell and Karine Harutyunyan
# Contents

About the Series ................................................................................................. v  
Abstract .................................................................................................................. vii  
About the Authors ................................................................................................. ix  
Acknowledgments .................................................................................................. xi  
Abbreviations and Acronyms ........................................................................... xiii  
Executive Summary ............................................................................................. xv  
Introduction ........................................................................................................... 1  
Overview of the Student Assessment System ................................................... 5  
Classroom Assessment ......................................................................................... 10  
Examinations .......................................................................................................... 15  
Large-Scale Assessments ..................................................................................... 26  
Discussion and Lessons Learned ......................................................................... 32  
Conclusion ............................................................................................................. 40  
References ............................................................................................................. 41  

## Boxes

Box 1: Overview of the Grades 4, 9, and 12 state graduation examinations .......... 17  
Box 2: Overview of the Unified Examination ....................................................... 20  
Box 3: Overview of the status, functions, and resources of the ATC .......... 22  

## Figures

Figure 1: Structure of the Armenian Education System as of 2015 ............... 4  
Figure 2: Annual enrolment data for the final year of general secondary education (G11 in 2010/2011 and G12 thereafter)............. 39
About the Series

Building strong education systems that promote learning is fundamental to development and economic growth. Over the past few years, as developing countries have succeeded in building more classrooms, and getting millions more children into school, the education community has begun to actively embrace the vision of measurable learning for all children in school. However, learning depends not only on resources invested in the school system, but also on the quality of the policies and institutions that enable their use and on how well the policies are implemented.

In 2011, the World Bank Group launched Education Sector Strategy 2020: Learning for All, which outlines an agenda for achieving “Learning for All” in the developing world over the next decade. To support implementation of the strategy, the World Bank commenced a multi-year program to support countries in systematically examining and strengthening the performance of their education systems. This evidence-based initiative, called SABER (Systems Approach for Better Education Results), is building a toolkit of diagnostics for examining education systems and their component policy domains against global standards, best practices, and in comparison with the policies and practices of countries around the world. By leveraging this global knowledge, SABER fills a gap in the availability of data and evidence on what matters most to improve the quality of education and achievement of better results.

SABER-Student Assessment, one of the systems examined within the SABER program, has developed tools to analyze and benchmark student assessment policies and systems around the world, with the goal of promoting stronger assessment systems that contribute to improved education quality and learning for all. To help explore the state of knowledge in the area, the SABER-Student Assessment team invited leading academics, assessment experts, and practitioners from developing and industrialized countries to come together to discuss assessment issues relevant for improving education quality and learning outcomes. The papers and case studies on student assessment in this series are the result of those conversations and the underlying research. Prior to publication, all of the papers benefited from a rigorous review.
process, which included comments from World Bank staff, academics, development practitioners, and country assessment experts.

All SABER-Student Assessment papers in this series were made possible by support from the Russia Education Aid for Development (READ) Trust Fund. READ Trust Fund is a collaboration between the Russian Federation and the World Bank that supports the improvement of student learning outcomes in low-income countries through the development of robust student assessment systems.

Papers in this series represent the independent views of the authors.
Abstract

In order to maximize effectiveness, every national education system needs a comprehensive and coherent approach to student assessment. The SABER framework for student assessment identifies four key elements in this regard: classroom assessment, examinations, national large-scale assessment, and international large-scale assessment. Armenia has undoubtedly made significant progress in all of these areas since gaining independence from the Soviet Union in 1991. The purpose of this paper is to describe Armenia’s experiences, share the lessons learned along the way, and point out the challenges that countries face in trying to ensure that their assessment systems remain fit for purpose in a changing environment.
About the Authors

George Bethell is an independent consultant specializing in the policy and practices of educational assessment. Having started his career as a physics teacher in a state secondary school in the United Kingdom, he subsequently served as a subject specialist at the University of Cambridge schools’ examination board where, amongst other things, he was responsible for the preparation and scoring of science examination papers, and for standards setting. Having become involved in international assessment whilst in Cambridge, George has, for more than 30 years, provided consultancy and training services to governments, international development agencies, and examining authorities around the world. His special interests include high-stake examinations and university admission systems, and sample-based surveys of student achievement. Since 1990, he has contributed to the reform of traditional assessment practices in more than a dozen of the transition countries of Southern, Central, and Eastern Europe. He has co-authored articles documenting major assessment reforms in Slovenia, Latvia, Lithuania, and Armenia. From 2005 to 2010, he led the technical team that supported the successful implementation of a centralized university admissions examination in Armenia as well as the establishment of the Assessment and Testing Center, which now serves as Armenia’s executive agency for school examinations and large-scale assessments of student achievement.

Karine Harutyunyan is a Deputy Minister in the Armenian Ministry of Education and Science (MOES). She holds a PhD in Physics and Mathematics, and a Master’s degree in Social Sciences from Yerevan State University where, in parallel with her responsibilities in the MOES, she teaches and does research. She has published in the region of 30 articles and books. Formerly, Karine served as Director of the Center for Education Projects, overseeing projects in the general and higher education sectors, including those supported by the World Bank. In this role she was responsible for the implementation of the first Education Quality and Relevance Project, which, as described in this case study, included components designed to reform classroom assessment practices, introduce a centralized examination at the interface between schools and institutions of higher education, and establish the national ATC. She has
also served as an independent consultant for, amongst others, the World Bank, the Asian Development Bank, and UNESCO. Her main areas of expertise are education policy, analysis and strategic planning, the financing of education, and project design and evaluation.
Acknowledgments

The authors are grateful to the READ Trust Fund team for the opportunity to present Armenia’s experience in assessing student achievement as part of the SABER Working Paper series.

Thanks are also extended to Arsen Baghdasaryan, Deputy Director of the Assessment and Testing Center and National Research Coordinator for TIMSS in Armenia for his generous collaboration.

Special thanks are due to the World Bank review team, Marguerite Clarke, Chelsea Coffin, and Cristian Aedo for their insightful suggestions on improving the case study and their close attention to detail throughout the editorial process.
### Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD</td>
<td>Armenian Dram</td>
</tr>
<tr>
<td>ATC</td>
<td>Assessment and Testing Center</td>
</tr>
<tr>
<td>CFEP</td>
<td>Center for Education Projects</td>
</tr>
<tr>
<td>CPD</td>
<td>Continuous Professional Development</td>
</tr>
<tr>
<td>EQRPI</td>
<td>First Education Quality and Relevance Project</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GORA</td>
<td>Government of the Republic of Armenia</td>
</tr>
<tr>
<td>HEI</td>
<td>Higher Education Institutions</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communications Technology</td>
</tr>
<tr>
<td>IEA</td>
<td>International Association for the Evaluation of Educational Achievement</td>
</tr>
<tr>
<td>ILSA</td>
<td>International Large-Scale Assessment</td>
</tr>
<tr>
<td>MOES</td>
<td>Ministry of Education and Science</td>
</tr>
<tr>
<td>NAEP</td>
<td>National Assessment of Educational Progress</td>
</tr>
<tr>
<td>NCET</td>
<td>National Center for Educational Technologies</td>
</tr>
<tr>
<td>NIE</td>
<td>National Institute for Education</td>
</tr>
<tr>
<td>NLSA</td>
<td>National Large-Scale Assessment</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Cooperation and Development</td>
</tr>
<tr>
<td>PIRLS</td>
<td>Progress in International Reading Literacy Study</td>
</tr>
<tr>
<td>PISA</td>
<td>Program for International Student Assessment</td>
</tr>
<tr>
<td>READ</td>
<td>Russia Education Aid for Development</td>
</tr>
<tr>
<td>SABER</td>
<td>Systems Approach for Better Education Results</td>
</tr>
<tr>
<td>TIMSS</td>
<td>Trends in International Mathematics and Science Study</td>
</tr>
<tr>
<td>UE</td>
<td>Unified Examination</td>
</tr>
<tr>
<td>USD</td>
<td>United States Dollar</td>
</tr>
<tr>
<td>USSR</td>
<td>Union of Soviet Socialist Republics</td>
</tr>
<tr>
<td>VET</td>
<td>Vocational Education and Training</td>
</tr>
</tbody>
</table>
Executive Summary

In 1991, Armenia emerged from the breakup of the Union of Soviet Socialist Republics (USSR) as a small, relatively poor, independent state facing significant socio-economic challenges. Since that time, it is widely acknowledged that the country has successfully built a comprehensive and coherent student assessment system to complement wider reforms in education. This case study describes how legacy practices from the Soviet period were transformed.

Progress has been made in the areas of classroom assessment, examinations, and large-scale surveys of student achievement at the national and international levels. At the center of this has been the establishment of a fully-functioning Assessment and Testing Centre (ATC) to serve as the national center of excellence for educational assessment. Primarily established to develop and administer a new, high-stakes Unified Examination (UE) at the school/university interface, the ATC has gradually assumed responsibility for other school examinations, national assessments, and Armenia’s participation in the Trends in International Mathematics and Science Study (TIMSS).

In the field of classroom assessment, the lead agencies are the Ministry of Education and Science (MOES) and the National Institute for Education (NIE). Building on activities to introduce formative assessment practices in schools carried out under the World Bank-supported First Education Quality and Relevance Project (EQRPI), the MOES has issued teacher-friendly regulations on classroom assessment, with teachers being further supported through methodological guidelines prepared by the NIE.

This case study discusses the drivers and mechanisms involved in pushing forward on so many fronts. The key driver was and remains the Government’s commitment to improving education in order to strengthen the country’s most valuable economic resource—its people. This can be seen in its willingness to invest heavily, not just in infrastructure, but also in the ‘softer’ areas of education, such as the development of modern subject curricula, learning standards, and assessment practices. External support and advice from the World Bank added momentum and ensured that no aspect of assessment was
neglected in the design and implementation of projects facilitated by credit agreements.

With respect to the development of the UE, a second driver commonly found in former-USSR countries added momentum—the desire to fight corruption in the university admissions system. In short, the introduction of the UE was seen as an opportunity to increase fairness and transparency and to reduce malpractice.

Finally, the case study highlights three key lessons learned whilst modernizing assessment practices in Armenia, which are likely to be of value to those still on the road to reform. These are:

- The technical elements of new assessments are relatively easy to introduce; it is far more difficult to ensure that the information they generate is used effectively. Assessment specialists must take the lead in ensuring that policy makers understand the potential power of assessment-related data. They also have an important role to play in ensuring that teachers get the information they need, in formats they understand, in order to improve teaching and learning.

- In the case of examinations, where the stakes are high and the public perception of corruption is great, there is inevitably a tension between the need to maintain examination security and the desire to promote important educational objectives through more innovative and effective approaches to examination design. Achieving an acceptable and effective compromise is a constant challenge.

- Assessments are not isolated from other aspects of society. As policies, priorities, and other contextual factors change, there will be a continual need to review the ‘fitness for purpose’ of assessment practices and, where necessary, adapt them. In short, in student assessment, as elsewhere, the reform process is continuous and without end.
The Development of the Student Assessment System in the Republic of Armenia: Achievements, Challenges, and Lessons Learned

George Bethell and Karine Harutyunyan

Introduction

In order to be effective, a nation’s education system must incorporate technically sound student assessment practices serving a wide range of purposes. These purposes include not only the certification of student achievement (i.e., assessment of learning), but also the promotion of better teaching and learning (i.e., assessment for and as learning) and the provision of objectively-verifiable evidence to inform policy making and strategic planning (i.e., system monitoring and evaluation). There is no universal blueprint for building such a comprehensive assessment system because the best-fit solution for a particular country will depend upon its objectives and priorities, geopolitical context and traditions, and the fiscal and human resources available. However, governments aiming to strengthen assessment practices can learn important general lessons from those that are further down the road to reform. The main purposes of this case study are to document the progress made by Armenia in the area of assessment over the past two decades and to share the lessons learned with other countries aiming to develop their own assessment systems.

This paper considers three types of assessment activities in Armenia’s general education system: classroom assessments, examinations, and large-scale assessments (both national and international). Innovations and modifications to pre-existing assessment
practices are described, and an attempt is made to identify the primary drivers that provided the momentum for change and the mechanisms that led to successful reforms.

Armenia is a particularly interesting case because it is often cited as an example of a relatively poor country that has made significant and rapid progress in the reform of student assessment practices. (See, for example, the SABER-Student Assessment Country Report for Armenia, 2011.) Its achievements, however, can only be fully appreciated by understanding the context in which reforms were undertaken.

Armenia was officially recognized as an independent country in September 1991, after the breakup of the Union of Soviet Socialist Republics (USSR). It is a relatively small country in challenging geographical and political environments.\(^1\) The population is just 3 million, but, as a result of turbulent events in the early twentieth century, there is a diaspora of approximately 8 million ethnic Armenians living outside of the country. Compared with other former socialist republics, the internal population is very homogenous, comprising 98 percent ethnic Armenians. There is a strong sense of national identity that is reinforced through a national language spoken only by Armenians and written in a unique script.\(^2\)

The country’s economic fortunes have fluctuated dramatically since independence. Immediately following the transition, the economy was weak and heavily dependent upon remittances from the diaspora. However, strategic decisions by the Government of the Republic of Armenia (GORA) generated impressive growth over a 15-year period.\(^3\) Unfortunately, the economy was hard hit by the global financial crisis of 2008, with a 14 percent drop in Gross Domestic Product (GDP) being recorded in 2009. There are now signs of a recovery but the current GDP

---

1. Armenia is a landlocked country in the mountains of the South Caucasus region sharing borders with Turkey, Georgia, Iran, and Azerbaijan. It has an area of 29,743 km\(^2\).
2. The Armenian alphabet is unique and, with the exception of a few later additions, remains unchanged from that introduced by the theologian and scholar Mesrop Mashtots around 405 Common Era.
annual growth rate of between 3 and 5 percent remains far below the double-digit growth of the pre-crisis period.

In Armenia, education has long been regarded as the main factor in maintaining the nation’s identity, survival, and progress, and, as a result, remains highly valued by the public and a priority for the Government. Since independence, significant reforms have been implemented across all aspects of the education system. Some vestiges of the Soviet system have been retained, but only those considered beneficial. The structure of the education system, including both general and professional education programs, is shown in Figure 1. General education for approximately 370,000 students is provided by 1,388 state schools and 47 private schools. Just 2 percent (~7,500) of the cohort attend private schools. In addition, approximately 36,000 study in Vocational Education and Training (VET) colleges (99) or schools (25). There are 70 Higher Education Institutions (HEI), of which 22 are state-owned, 35 are privately-owned, five are intergovernmental, and eight are branches of foreign universities. 4 Approximately 103,000 students are currently enrolled in HEI.

This paper discusses three main types of assessment activity in Armenia’s general education system: classroom assessments (particularly those designed to yield information to support teaching and learning), examinations (where results are used to make high-stakes decisions about individual students), and large-scale assessments (for monitoring performance at the system level). 5 For each assessment type, three key factors (i.e., ‘quality drivers’) are considered. These are: the enabling context in which assessment activities take place; the degree to which assessment activities are aligned with broader aspects of the education system, including, for example, curricular objectives and goals; and the technical quality of the assessment, including its validity with respect to its intended purposes.

---

4 Intergovernmental HEI in Armenia include the Russian-Armenian (Slavonic) University, the American University of Armenia, the French University of Armenia, the European Regional Educational Academy of Armenia, and the French Higher Institute of Engineering in Armenia.

5 This reflects the general structure of the READ/SABER Framework for Student Assessment Systems, as described in Clarke (2012).
This general introduction is followed by an overview of the current student assessment system in Armenia. Thereafter, the three types of assessment are discussed in detail. For each, the paper provides a description of current practices, policies, structures, and resources. It also identifies and evaluates key drivers and mechanisms of change. The final section of the paper summarizes important issues and findings. In addition, key lessons learned—especially those that may be of interest to others reforming or strengthening their student assessment systems—are presented.
Overview of the Student Assessment System

The current student assessment system in Armenia is comprehensive, both in its coverage of the primary and secondary phases of general education and in the variety of assessment practices used. The system may have some redundancy or inefficiencies, and there are undoubtedly areas where further reforms are needed, but the fundamental structure is sound, as reflected in the generally-positive findings of the SABER-Student Assessment benchmarking exercise conducted in 2011 (World Bank, 2011). This is a far cry from the situation inherited from the Soviet period, the major weaknesses of which are described below.

Before the launch of reforms, the Ministry of Education and Science (MOES) was responsible for all forms of student assessment, including state graduation examinations. Unfortunately, but not surprisingly, the Ministry was not well placed to manage this huge task. First, it lacked the necessary human, physical, and financial resources. For example, during examination sessions, almost all Ministry staff were involved in logistics and administration, meaning that all other ministerial functions were effectively paralyzed. Second, but equally important, the MOES was hampered by a narrow view of assessment as ‘testing’, which was reinforced by limited expertise in this specialized field.

Continuous classroom assessment was left entirely to the school and classroom levels. Teachers assiduously followed administrative guidelines on the testing of pupils, but feedback mechanisms were at best weak. The assumption was that all teachers were reliable assessors, but the truth is that they were poorly prepared, with little or no training in assessment techniques and their uses.

For the public, the greatest concerns were reserved for the university entrance examinations conducted by subject commissions in universities. Formally, the examination papers were supposed to be based on the upper-secondary school curriculum. In practice, however, there was a significant gap between what was taught in schools and the requirements for admission. This placed extra pressure on students preparing for entrance examinations and led most parents to believe that expensive private tutoring was essential for the success of their children. In addition, a number of special arrangements made the system inequitable. For example, some specialized upper-secondary schools reached
agreements with university departments that exempted their students from entrance examinations and required them only to pass an interview. Similarly, students who had won a ‘Gold Medal’ or had been successful in a subject Olympiad had to take only one examination, with some being admitted simply on the basis of an interview. In short, prior to the reforms, the traditional system of university entrance examinations was inefficient and unfair.

Through policies adopted by the Government, technical development activities undertaken through World Bank-supported projects, and regulations and guidelines enacted by the MOES, many of these issues have been addressed, as seen in the descriptions of current assessment practices which follow.

**Classroom Assessment**

The continuous assessment of students by teachers was a significant feature of the Soviet system and one that was valued and understood, not only by teachers, but also by students and their parents. As in many other former Soviet republics, Armenia has retained desirable elements of this school-based system, whilst adjusting other aspects to correspond to current thinking and pedagogical objectives. For example, in line with traditional practices, teachers are encouraged to use a range of familiar assessment techniques, the most common of which are ‘personal enquiry’ (i.e., questioning and oral tests), ‘control works’ (i.e., written tasks set specifically for the purpose of assessment), and assessed homework tasks. Teachers grade all forms of assessment on a ten-point scale, where 9 to 10 is ‘excellent’, 7 to 8 is ‘good’, 5 to 6 is ‘satisfactory’, and, 1 to 4 is ‘failing’. Student results are recorded in the class journal. There is no formal standardization across schools, but the scoring system is familiar to all and scores are widely used as sufficiently reliable indicators of achievement. In short, the educational authorities and the public continue to place great trust in the evaluations made by schools and teachers. It should be noted, however, that the primary function of these assessment activities is to record student achievement; i.e., they are summative in nature.

---

6 For example, schools with specializations in mathematics and physics had this agreement with the Mathematics and Physics Departments of Yerevan State University.
Notwithstanding the above, significant efforts have been made to re-direct classroom assessment practices towards gathering information to be used diagnostically in support of ongoing teaching and learning. For example, the World Bank-supported First Education Quality and Relevance Project (EQRP I) implemented an extensive in-service training exercise to introduce teachers to the use of assessment techniques for purely formative purposes. This approach is required by the government-approved Student Assessment Concept (GORA, 2005) document and supported by methodological guidelines for teachers published by the National Institute for Education (NIE, 2013). It should also be noted that teacher assessment in Grade 1 is designated for diagnostic and monitoring purposes only—summative assessment in the first year of schooling is explicitly prohibited by the Law on General Education (GORA, 2009).7

It would be inaccurate to say that all teachers are now fully committed to the concept of formative assessment, but the degree of ‘professional inertia’ is probably no greater than that encountered in many other countries, including those with far greater resources than Armenia.8

Examinations

State graduation examinations at the ends of Grades 4, 9, and 12 are a significant feature of the assessment system. The Grade 4 examination is intended to certify that students have successfully completed the primary phase of education, and focuses on Armenian Language and Mathematics. Individual schools set their own tests according to guidelines approved by the MOES and mark their own students’ papers. Very few students fail this examination—in 2012, just 33 students failed out of 3,1254—and there is a growing awareness that this legacy examination now serves little purpose since transition to Grade 5 is, to all intents and purposes, automatic.

---

7 Article 16, para. 3 states that “Examination of knowledge of a child during … the first year of general education school shall be prohibited.”

8 For example, a report into the impact of a national strategy for supporting assessment for learning in schools in England reported that “although teachers and senior leaders valued the training and support they had received … this did not necessarily lead to effective assessment for learning in their schools” (Ofsted, 2008, p.4).
The Grades 9 and 12 state graduation examinations cover five subject areas: Armenian Language and Literature, Mathematics, Armenian History, Sciences, and Foreign Languages. The question papers are prepared by the Assessment and Testing Centre (ATC) and administered in schools by an examination commission established by the school’s Director. Pass rates at Grades 9 and 12 are extremely high (in excess of 98 percent). The examinations have little power of discrimination and merely serve to indicate that a student has completed the relevant phase of education and has reached very minimal levels of achievement. Success is recorded on a student’s school completion certificate (diploma), which is a necessary qualification for continuing education in, for example, VET institutions and also has some value in the employment market.

For school leavers wishing to enter higher education, the Unified Examination (UE), introduced in 2006, serves two purposes: school graduation and university entrance. As such, it tends to dominate the secondary phase of education and the stakes associated with it are extremely high. The examination is developed and conducted under the auspices of the ATC. The UE exerts considerable pressures on schools, teachers, students, and their families and this tends to distort not only what is taught, but also how it is taught. In particular, many candidates receive additional private tutoring to increase their chances of success on the UE. Notwithstanding any negative impact that it may have, the UE represents a marked improvement on the Soviet system of decentralized university entrance examinations which it replaced. In particular, it is demonstrably more equitable, more transparent, and less prone to corruption.

**National and International Large-Scale Assessments**

Classroom assessments and examinations focus on measuring the achievement of individual students for specific purposes. They are not, however, well designed for monitoring overall standards at the system level, or for investigating the background factors that can contribute to higher levels of student achievement. These functions are better served by large-scale assessments conducted at the national and international levels. Since 2003, Armenia has participated in the Trends in International
Mathematics and Science Study (TIMSS). It is currently participating in the 2015 cycle.⁹

More recently, in order to complement its involvement in TIMSS, Armenia has designed its own national, large-scale assessment (NLSA). This is a sample-based survey that adopts several of the technical procedures of TIMSS, but focuses on achievement in key subject areas of the Armenian curriculum. The ATC is responsible for all aspects of the national assessment and conducted the first cycle in 2010. The 2010 assessment focused on student achievement in the subject areas of Armenian Language and Literature, and Armenian History.

Policies, Organizational Structures, and Resources

Key policies concerning student assessment procedures are set out in government legislation. The primary legislation is the Law on Education adopted in 1999 (GORA, 1999). Subsequent laws regulate the separate levels and functions of the education sector. Most relevant here is the Law on General Education of 2009 and subsequent amendments (GORA, 2009).

The MOES oversees the implementation of the law in schools and as such is responsible for setting regulations for the delivery of the curriculum and the assessment of learners. Norms for classroom assessments and the organization of the state graduation examinations are set by Ministerial decree. The most recent of these is the decree prescribing Procedures for Assessment of Student Progress, adopted in 2014 (MOES, 2014). The only form of assessment that does not fall directly under the auspices of the MOES is the UE. This is conducted by the ATC which, since 2009, has been subordinated to the Government Cabinet and hence the Prime Minister.¹⁰ Policies concerning the UE are set out in the prevailing laws and amendments on General Education.

The ATC serves as the lead institution for examinations and large-scale assessments. There are 81 members of staff, including 15 assessment professionals in the Department of Testology and 12

---

⁹ In Armenia, the main data collection for TIMSS 2015 was completed in May 2015. The international report is due to be released in December, 2016.

¹⁰ The ATC was formally created on April 14, 2004 by Government Decision #586-N as a separate institution under supervision of the MOES, with the legal status of a state, non-profit organization. It was transferred to the Government Cabinet in 2009.
specialists in the Department of ICT and Analysis. ATC is funded through a dedicated line in the State budget. The 2014 budget allocated a total of AMD 413 million (~USD 1 million) for annual costs, including staff salaries and utilities. ATC is also permitted to generate income through student fees for taking the UE. In 2013, this generated approximately AMD 36 million (~USD 90,000) in income.

Beyond budgetary provision for ATC, few, if any, funds are explicitly ear-marked for student assessment. In particular, teachers are expected to conduct continuous assessment and mark state graduation examinations as part of their normal duties, i.e., without extra payment. Similarly, any costs incurred in organizing graduation exams are met from school budgets.

**Classroom Assessment**

**Overview**

At independence, Armenia inherited a robust system of classroom assessment in which teachers were required to evaluate and record student achievement regularly so that progress could be monitored by parents and the educational authorities. All forms of assessment were graded on a ubiquitous five-point scale, which was universally understood even though there was no standardization within or between schools.\(^\text{11}\) Indeed, teachers enjoyed—and continue to enjoy—a great deal of public trust in their role as assessors. This legacy formed a firm foundation on which to build modern reforms, including promoting the use of classroom assessments for formative purposes.

Current MOES regulations (MOES, 2014) are in place for a comprehensive system of school-based assessment, which requires teachers to assess their students regularly, using a variety of methods, for both formative and summative purposes and to record and report

\(^{11}\) On this traditional scale, 1 and 2 were considered to be failing grades, 3 indicated ‘satisfactory’, 4 was ‘good’, and 5 was ‘excellent’. This scale has now been replaced by a ten-point scale, but the general categories remain unchanged.
outcomes on a standard ten-point scale.\textsuperscript{12} This is well understood by schools and, from Grade 2 onwards, students are assessed through oral tests, assessed tasks (‘control works’), and occasional, written tests. In addition, more formal, end-of-semester tests in each curriculum subject are arranged by schools. Grades (on a ten-point scale) are assiduously recorded and used to calculate grade-point averages for the year. In practice, the balance is still heavily in favor of summative assessment, with ‘pure’ formative assessment found in a minority of classrooms. Part of this is due to the fact that, in spite of in-service training, many teachers remain unfamiliar with diagnostic assessment techniques. In addition, the pressure of competition means that schools still feel compelled to measure and report student progress to parents in more traditional ways.

Initial MOES guidance to teachers, designed to regularize and control classroom assessments in accordance with the practices developed under EQRPI, drew many complaints. Teachers argued, with some justification, that the procedures were overly complicated and that they increased their workload unacceptably. In response, the new guidelines offer more clarity and greater freedom for teachers to choose how often, when, and how they use summative and formative assessments in their classes.\textsuperscript{13} They also provide clear and simple guidance on the aggregation of scores from summative assessments to produce annual grades.

Teachers are further supported by a practical guide produced by the NIE (2013). In addition, further initiatives are being introduced to help teachers overcome their lack of confidence in applying new classroom assessment practices. These are described under \textit{Mechanisms of Change} below.

**Current Policies, Structures, and Resources for Classroom Assessment**

GORA and the MOES have set out clear policies concerning the continuous assessment of students by teachers. Building on a Ministerial Decree, ‘Concept of Students’ Continuous Assessment: General

\textsuperscript{12} The MOES issued new guidelines on classroom assessment in 2014 to replace earlier guidelines that had drawn vociferous objections from teachers as being too complicated and adding to their work burden.

\textsuperscript{13} The guidelines do set minimum numbers for each type of assessment to ensure that teachers use an appropriate variety of assessment techniques.
Statement, Basic Approaches’ (MOES, 2008), current requirements are set out in the MOES guidelines, ‘Procedures for assessment of student progress’ (MOES, 2014). The MOES, in close collaboration with its schools, monitors the implementation of classroom assessment policies. Additional professional support is offered through the NIE.

Few resources are explicitly allocated to classroom assessment since it is seen as an integral part of the teaching/learning process which, in turn, is funded through school budgets.

**Mechanisms of Change**

EQRP I had a broad scope, including the development of new curricula and standards and harnessing the positive backwash effect associated with high-quality assessment in schools. The project allowed the MOES to launch, in 2005, a series of initiatives to develop a system of continuous, formative assessment in classrooms, with the explicit objective of encouraging schools and teachers to use assessment results for diagnostic purposes. The defining concept was produced by a specialist working group and then suggested methods were validated through extensive consultation and piloting (MOES, 2008). Finally, an experimental, three-year program was approved by the MOES and implemented in all schools.\(^\text{14}\) Part of the implementation strategy was to provide schools with model tests in a wide range of subjects, based on new curricula and standards developed under EQRP I. Model tests and guidance for their use were distributed to all schools “so that teachers could learn new questioning techniques that elicit better information and lead to the identification of any potential ‘gap’ between what has been taught and what has been learned” (Center for Education Projects [CFEP] 2009, p.24).

Implementation was supported by a sub-component of EQRP I designed to develop a short, modular program for the in-service training of teachers in the principles and practices of assessment. A working group of central trainers worked with international consultants to design the modules, the materials, and a training program for local trainers who, in turn, were to train practicing teachers throughout the country. Whilst

this round of in-service training was generally well-received, its long-term impact is difficult to judge. More importantly, newly-qualified teachers joining the profession do not have the benefit of having received intensive training in classroom assessment methods. Therefore, MOES, with the support of the Russia Education Aid for Development (READ) Trust Fund, is currently working on new mechanisms to build upon the progress made thus far.

The first initiative is a small, but highly significant, part of a far more fundamental change related to the recent introduction of a well-defined career structure for teachers. This gives teachers the opportunity to improve their professional status—and their salaries—by progressing through a series of four levels or ranks. Progression is dependent upon a number of factors, including years of service and participation in a formal system of Continuous Professional Development (CPD). The CPD element requires teachers to take at least 80 hours of training over a five-year period. The fields in which training is required include assessment and so, through this mechanism, teachers will be able to access in-service training from recognized providers, the most predominant being the NIE.

The second initiative is the development of a pre-service course for teachers to ensure that new entrants to the profession have a sound understanding of assessment principles and practices. The READ Trust Fund is supporting the implementation of this initiative by providing technical assistance to the seven HEI that prepare teachers so that an appropriately comprehensive and demanding course can be developed. The course will likely place an emphasis on modern approaches to formative assessment and its use in the classroom (Boyle, 2014).

The third initiative, which should in the longer-term strengthen the professional assessment community in Armenia, is the development of a two-year Master’s course in Assessment, Measurement, and Testing. This is in the early stages of development and is being supported with technical assistance provided by the READ Trust Fund. Classroom assessment methods will likely be one of the options available to students for in-depth study and research (Hawker, 2013).
Drivers of Change

At the time when EQRP I was being prepared, there was growing awareness in the international education community that assessment for learning (formative assessment) had the potential to deliver improved educational outcomes for relatively little cost. In particular, the ideas that Black and Wiliam (1990) described in their seminal work, "Inside the Black Box: Raising Standards through Classroom Assessment," were increasingly being accepted. In Armenia, it was recognized that the conditions were right for including formative assessment in reform activities because the country had maintained from Soviet times a respect for the teacher's role in evaluating student progress. The existence of a professional teaching force with a tradition of compliance with MOES requirements provided an ideal enabling context for change.

Once included in the project, program funding and the active management of the project implementation unit (Center for Education Projects [CFEP]) served as additional drivers. The staff and consultants of the CFEP developed excellent working relationships with members of specialist working groups. This was particularly important in the development and implementation of the in-service training program for teachers.

Another important driver of change was the emergence of a new assessment community as a direct result of activities conducted under EQRP I. For example, many subject specialists who had been involved in the development of model question papers for the new, high-stakes UE continued their involvement as members of the working group that developed the training program for classroom assessment. This continuity of engagement allowed them to develop a deeper understanding of the fundamental principles that underpin good assessment practices.15

15 The current Director of the ATC was trained in the principles of assessment as a member of an EQRP I subject working group.
Examinations

Overview
Within the general education system, students encounter two types of formal examination: state graduation examinations at the ends of Grades 4, 9, and 12; and the UE, which sits at the interface between secondary and tertiary education.

The general system of state graduation examinations is a legacy from Soviet times, and the main purpose is to certify the successful completion of primary, middle-secondary, and upper-secondary schooling. However, the overall pass rates for these examinations (after permitted retakes are taken into account) approach 100 percent. As a result, they do not present a major hurdle to students in continuing their education and hence the stakes are relatively low. Indeed, some have argued that the Grade 4 graduation examination in particular is now redundant. However, others hold that the established pattern of formal examinations does serve to punctuate the teaching and learning process for students, and the Grades 9 and 12 graduation exams continue to have some, albeit limited, concrete value.

By way of contrast, the stakes associated with the UE taken at the end of Grade 12 are extremely high. As a result, the content of the examination influences, and some would say distorts, not only what is taught, but also how it is learnt (READ, 2012). The examination, its impact on teaching and learning, and plans for its future development are discussed further below.

State graduation examinations
The general system of state graduation examinations was inherited from Soviet times. However, significant changes were made to the design and delivery of the examinations during the period of reform. First, the grade levels were adjusted to match the longer 12-year period of general education, and the grading scales were changed to match the 10- and 20-point scales introduced under the new Concept of Assessment, adopted in 2005 (GORA, 2005). Second, the content, format, and mode of delivery of the tests were changed. For example, in the Soviet period, examination tasks were prepared by the NIE with questions and
problems selected from textbooks. Then, on the specified day and time, these were transmitted by radio. Most of the items on the tests, with the exception of those for Armenian and foreign languages, were open-ended and scored subjectively by teachers (Bethell and Harutyunyan, 2008). Now, the tests contain only items that can be scored objectively, i.e., multiple-choice and short-answer types.

State graduation examinations focus on the key subjects of the school program and test papers are closely aligned with the content of the national curriculum and standards. Box 1 summarizes the main characteristics of the three examinations. It can be seen that responsibility for test administration, the marking of students’ work, and the issuance of results is effectively delegated to schools. Steps are taken to ensure that schools follow key administrative guidelines, test papers are kept secure, and marking is fair. Indeed, schools take the examinations seriously and the results are, in general, trusted by students, parents, and the wider public.

A significant change in the administration of the Grades 9 and 12 examinations was made possible by the establishment of the ATC under EQRPI. Originally, the ATC’s priority was the design and conduct of the UE. However, once it was operational, the Ministry was able to switch responsibility for the production of the Grades 9 and 12 test papers to the ATC, thereby standardizing tests across schools. In the past, little quantitative information was gathered (except for school failure rates) and no secondary analysis was carried out. However, in 2013, for the first time, ATC analyzed the results of the graduation examinations for Grades 9 and 12 and published reports on its findings. This is a positive step, showing ATC’s desire to strengthen the use of assessment-related data and improve feedback systems.
Box 1: Overview of the Grades 4, 9, and 12 state graduation examinations

<table>
<thead>
<tr>
<th>Grade 4 State Graduation Examination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjects</strong></td>
<td>Two subjects: Armenian Language and Mathematics (90 minutes each). An additional Physical Training examination is also arranged.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>Guidelines for the examinations are set by MOES and the ATC provides sample/model examination papers. The school Director is responsible for implementation. The school’s subject specialists set five versions of the tests. The five versions are distributed spirally to prevent copying. Teachers mark their students’ work and report results on a 10-point scale: 10/9 = 'excellent', 8/7 = 'good', 6/5 = 'satisfactory', and 4-1 = 'fail'. Students who fail are allowed to retake the examination.</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>After permitted retakes, the pass rate in 2011 was 99.9 percent—only 33 students failed out of 31,254.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 9 State Graduation Examination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjects</strong></td>
<td>Five subjects: Armenian Language and Literature, Mathematics, Armenian History, a Science (Physics, Chemistry, Biology, or Geography), and Foreign Language (typically Russian, English, French, or German) (90 minutes each). An additional Physical Training examination is also arranged.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>Test papers and marking criteria are set by the ATC. Tests are delivered to schools just two hours before the examinations are due to start. All other processes are conducted under the auspices of an examination commission composed of selected teachers and headed by the school’s Director. Students’ work is marked by the members of the commission, with results reported on 20-point scale.</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>Pass rates are close to 100 percent. (In 2011/2012, the pass rate was 99.6 percent, with about 130 students failing from a cohort of more than 38,000.) The Grade 9 diploma has some value in that it is necessary for continuing education in high schools or in VET institutions. For those leaving school, the diploma has some value in the employment market.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 12 State Graduation Examination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjects</strong></td>
<td>The subjects tested are the same as for the Grade 9 examination: Armenian Language and Literature, Mathematics, Armenian History, a Science (all 120 minutes each), and a Foreign Language (90 minutes). A Physical Training examination is also arranged.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>General arrangements as for the Grade 9 examination: ATC is responsible for delivering examination papers to schools in advance and the conduct of the examination and the marking/grading process fall under the auspices of an examination commission established and led by the school’s Director.</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>Typical pass rates for this examination (after retakes) are in excess of 99 percent. (The pass rate in 2012 was 99.15 percent—298 failed out of 35,244.) The Grade 12 diploma is important in that it is necessary for continuing education at the tertiary level in VET and HEI. For those leaving school, the diploma has some value in the employment market.</td>
</tr>
</tbody>
</table>
Unified Examination

For students wishing to apply for further studies in HEI, the Grade 12 UE serves as a school graduation test and, most importantly, the university selection examination. The latter role ensures that it dominates the secondary education system. Bethell and Hartyunyan (2008) describe the situation thus:

Throughout the world, examinations used to select applicants for institutions of higher education are said to have ‘high stakes.’ However, nowhere is this more true than in Armenia. Firstly, as in other countries, places on prestigious courses are prized for the economic and social advantages they bring. Secondly, Armenian students who do exceptionally well in selection tests are eligible for free tuition at university. Those who fall below the threshold score for state-funded places may still go on to study, but only if they can afford to fund themselves. The third factor, and by far the most important for young men, is that those who gain (state-funded) places at university may defer their compulsory military service (Bethell and Hartyunyan, 2008, p.112).16

The current system, in place since 2006, was specifically designed and implemented under EQRPI in order to replace the traditional Soviet system whereby students had to apply to, and meet the entry requirements of, individual universities (Bethell and Zabulonis, 2012). The main aims of the reform were to increase fairness and transparency within the selection process and to combat widespread perception of corruption and malfeasance in the old system. This required the establishment of an organization with the capacity and integrity necessary to conduct a high-stakes selection examination to the highest standards for the whole country. As a result, the ATC is now responsible for all aspects of the UE. This includes not only test production and scoring, but also the administration and logistics associated with conducting a high-stakes test under secure conditions.

In order to fulfil two distinct purposes—school graduation and selection—the UE has two tiers. Students who do not wish to use their results for a particular subject in their university application opt for level A, which examines the full subject curriculum, but only includes tasks up

---

16 Male applicants who gain state-funded places may defer national service until after their Bachelor studies. Those who gain fee-paying places have to enter the military immediately, but their places are held until they complete their two years of service.
to average levels of difficulty. Students wishing to use a particular subject in their university entrance application take an examination incorporating level A and the more-demanding level B. Applicants take the UE in special examination centers established by ATC. This centralization of test administration allows the examinations to be conducted in a highly-formal, controlled environment.

For graduation purposes, students take examinations in four compulsory subjects (Armenian Language, Mathematics, Armenian History, and one Foreign Language) and two elective subjects from a list approved by the MOES. However, students wishing to enter university also have to meet admissions criteria by taking examinations in the subjects required by the HEI courses to which they are applying. Applicants can apply to up to ten different universities. Some specialties and professions also require applicants to take additional, specialized examinations separate from the UE. By studying the list of examination subjects required by the universities and departments to which they have applied, an applicant can ascertain the total number of examinations that he or she has to pass under the UE. There is no limit to the number of examinations that a student can take from the government-approved list of subjects. However, candidates are required to pay a fee of AMD 1500 (~USD 3.60) for each subject.

For graduation purposes, a candidate’s raw score on the level A examination is transformed to a grade on the 1–20 scale, where 7.5 is the minimum passing grade. For university selection purposes, a candidate’s total raw score on levels A and B is converted to a score on the 20-point scale by a simple linear transformation. Here too, the minimum passing grade for all subjects is set at 7.5. This makes the procedure both transparent and capable of being explained to a highly-critical public, which would not be the case for more sophisticated non-linear transformations, e.g., normalization. The final score used to award free and paid places on university courses is calculated by the simple addition of an applicant’s scores in the combination of subjects required by the specific university and specialty. (See Box 2 for an overview of the UE system.)

---

17 Students taking the UE do not have to take the Grade 12 state graduation examination conducted in schools, since part A of the UE provides them with a score for their graduation diploma.

18 These costs are based on the 2014 ATC regulations for the UE.
Box 2: Overview of the Unified Examination

<table>
<thead>
<tr>
<th>Unified Examination and Supplementary Examinations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UE and other entrance examination subjects</strong></td>
</tr>
<tr>
<td>Subjects examined: Armenian Language and Literature, Armenian History, World History, Mathematics, Physics, Chemistry, Biology, Geography, Foreign Language (Russian, English, French, German, Spanish, Italian, or Persian).</td>
</tr>
<tr>
<td>Test format: Items capable of automated scoring through optical character recognition, i.e., multiple-choice and short (numerical)-answer questions (in Mathematics, Physics, and Geography).</td>
</tr>
<tr>
<td>Duration: Foreign Language examinations = 150 minutes; all others = 180 minutes.</td>
</tr>
<tr>
<td>Additional, centralized examinations are required in cases where speaking or writing skills have to be examined (e.g., for Journalism, Philology, Linguists, Foreign Languages, Literature, etc.). These are either ‘essays’ with a duration of 240 minutes, or oral tests (30 minutes per candidate).</td>
</tr>
<tr>
<td>For certain specialties, some universities require a specialized ‘Internal Examination’ in which candidates have to perform or demonstrate skills, e.g., Fine Arts, Music and the Performing Arts, Sports, Design, etc.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
</tr>
<tr>
<td>ATC is responsible for all aspects of the UE, including the preparation of test papers, organizing the conduct of the examination in approved centers, and the marking/grading process.</td>
</tr>
<tr>
<td>The main session for each subject examination is conducted on one day in 34 examination centers distributed all over the country. An additional session is organized for applicants who were registered, but absent on the day of the main examination for justified reasons.</td>
</tr>
<tr>
<td>The organization and conduct of the examinations requires enormous resources. In 2013, a total of 863 people were involved in logistics, organization, and supervision of the UE.</td>
</tr>
<tr>
<td>The biggest examinations are Armenian Language and Literature (8,647 candidates in 2013) and Mathematics (6,621). The smallest are in minority Foreign Languages, e.g., Spanish (20), Italian (4), and Persian (1 only)!</td>
</tr>
<tr>
<td><strong>Construction of UE test and score processing</strong></td>
</tr>
<tr>
<td>All UE tests contain 80 items (50 for Part A and 30 for Part B). Each item is scored 1 for correct answers or 0 for wrong answers, giving a maximum possible score of 80. An exception to this is the Armenian Language and Literature and Mathematics tests, which have some items that include six sequential sub-questions, each with three possible answers: ‘yes’, ‘no’, or ‘don’t know’. Each sub-question is scored 1 if correct, −1 if incorrect, or 0 if the chosen response is ‘don’t know’.</td>
</tr>
<tr>
<td>The total raw score (maximum 80) is transformed to the 1–20 scale simply by dividing the total by four. The minimum passing grade is set at 7.5.</td>
</tr>
<tr>
<td>In 2013, the overall failure rate was 28.4 percent of all test-takers.</td>
</tr>
<tr>
<td><strong>Examination fees</strong></td>
</tr>
<tr>
<td>Students registering for the UE are required to pay a fee of AMD 1500 for each subject (ATC 2014).</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
</tr>
<tr>
<td>The new system was phased in over several years. In 2007, only Armenian Language and Literature was examined. In 2008, Mathematics and four foreign languages were added. In 2009, the suite of subjects was completed.</td>
</tr>
<tr>
<td>The number of candidates has decreased significantly since 2007. For example, the number for Armenian Language and Literature has fallen from 14,682 to 8,934 (-39 percent) and for Mathematics, from 10,112 to 6,928 (-31 percent). The main reason for this is the decreasing school population.</td>
</tr>
</tbody>
</table>
Current Policies, Structures, and Resources for Examinations

The Law on General Education (GORA, 2009) sets out the general policy concerning end-of-phase school graduation examinations. In addition, the Law on Higher Education (GORA, 2004) establishes the principle that admission procedures to HEI at the Bachelor’s degree level “shall be defined by the Government of the Republic of Armenia”19, i.e., not by individual universities. In both cases, implementation of policy is delegated to the public bodies responsible—the MOES in the case of the state graduation examinations, and the ATC for the UE. In the case of UE, policies continue to evolve. First, in 2009, ATC was transferred to the direct subordination of the Government Cabinet/Prime Minister, thereby creating a degree of separation from the MOES. More recently, the ‘Program of Actions to Fight Corruption in the Education Sector 2011-2012’ proposed that question papers for the UE should be produced by the selection of items from a publicly-available, open item bank. This policy, implemented by ATC in 2012, is effectively a return to pre-2007 practices. One consequence of this is that all UE assessment tasks are now, by definition, set in familiar contexts, and hence all problem solving is routine. In short, using an open item bank reduces the opportunity to assess higher-level cognitive skills and promotes memorization and coaching for the test.20

State graduation examinations are delivered by the MOES (with some services provided by ATC) through school Directors and ad hoc school examination commissions. Few, if any, resources are explicitly dedicated to the preparation and conduct of the examinations. In particular, the administration of the examinations and the scoring of student responses are considered part of a teacher’s normal working contract and so no extra payment is made. By way of contrast, significant resources are allocated to carrying out the UE. From the outset (2005), ATC’s recurrent expenses have been funded directly through the state budget. Initially, EQRPI funds provided the capital expenditure necessary to refurbish suitable accommodation, and to provide ATC with

19 Article 14, paragraph 5.
20 The extensive use of private tutors is considered a problem in Armenia, as elsewhere, and it was hoped that the publication of item banks would reduce the pressure on parents to employ tutors. Unfortunately, anecdotal evidence suggests that coaching is still widespread and that tutors now train students simply to remember solutions rather than improving their subject-related knowledge and skills (Bagdasarayan, personal communication, 2014).
technical equipment, including an in-house printing facility and an IT system capable of processing the new, technologically-based examinations. When additional equipment needs were identified, the equipment was provided to ATC from the state budget. Through EQRPI, and with the continuous, active support of GORA, Armenia was provided with a “well functioning institution, fully refurbished and equipped with modern equipment… (with) nine administrative divisions and 70 staff members” (CFEP 2009, p.19). An overview of the ATC’s current status, functions, and resources are provided in Box 3.

Box 3: Overview of the status, functions, and resources of the ATC

| Status | Legal entity, established by charter, with the status of a state, non-profit organization. ATC is directly subordinated to the Government Cabinet and hence the Prime Minister. |
| Functions | The operational objectives of ATC are defined by its charter. These include the development and conduct of all forms of student assessment in the general phase of education. In addition, the ATC is charged with providing methodological guidance and training to teachers “in the efficient use of new forms of assessment” (GORA 2004, Appendix, cl. 12[4]). ATC is also required to conduct analyses of the outcomes of various assessments and to provide GORA with reports. Under its charter, ATC is required to “maintain close links with schools and HEI and to provide results of examinations (at the student level)” (ibid. Appendix, cl. 12[4]). Specific responsibilities currently include: |
| | • Conduct of the UE at the secondary/tertiary interface |
| | • Preparation of test papers and scoring criteria for the state graduation examinations for Grades 9 and 12 |
| | • Preparation of model test papers to guide schools in the preparation of the state graduation examinations for Grade 4 |
| | • Preparation and conduct of annual, sample-based national achievement surveys |
| | • Preparation and conduct of annual, sample-based ‘external evaluations’ of student achievement (Grades 5-11) |
| | • Conduct of ILSA (currently TIMSS) |
| Staffing | ATC has a total staff of 81 (ATC data 2014). The staffing structure is shown below. |
| | Senior Management (Director, Deputies, and Advisor) 4 |
| | Finance and Accounting 4 |
| | Legal and Staff Administration Department 5 |
| | Examination Organization Division 12 |
| | ‘Testology’ Division 15 |
| | External Relations 9 |
| | ICT and Analysis Department 12 |
| | Publishing 8 |
| | Logistics and auxiliary staff 12 |
| Financing | The 2014 state budget line for ATC allocated AMD 413 million (~USD 1 million) for annual running costs, including salaries and utilities. Within this, AMD 167 million (~USD 400,000) were for ‘contractual works’, i.e., the development, production, and conduct of examinations and assessments. ATC also generates income through student fees for taking the UE. For 2013, this is estimated at AMD ~36 million (~USD 90,000). |
Mechanisms of Change

The primary mechanism for building institutional capacity for the conduct of formal examinations in the general phase of education was EQRPI. Coordinating the wide range of project inputs associated with establishing the ATC, and simultaneously developing the UE model and instruments, fell to the Ministry’s CFEP. Key elements included civil works (refurbishing a building to provide accommodation for ATC) and the procurement of general and specialist equipment. Perhaps the most significant task was coordinating and checking the quality of services provided by international consultants responsible for developing human resources through the training of subject specialists as item writers and ATC professional staff. A summary of these activities and related outcomes can be found in the completion report for the project (CFEP, 2009).

Initially, advocates of the reforms met with cautious resistance from voices within the Government. In particular, the Ministry of Finance expressed grave concerns over proposals to establish a new, specialized institution at a time of general budgetary restraint, especially when student assessment had traditionally been conducted without any explicit support from the state budget. In order to overcome these understandable objections, the MOES and the CFEP had to prepare well-founded justifications set out in a range of formal documents. In addition, a series of meetings was held with relevant Ministries and the Cabinet of the GORA in order to explain the rationale behind the reforms and the potential long-term benefits. Without this advocacy, it is highly unlikely that ATC would have come into existence.

Once the ATC was formally established, identifying and recruiting the right personnel emerged as a major challenge. A Director was recruited with the necessary management skills, but there was a shortage of experienced assessment and IT/analytical specialists. To overcome this, people with the potential to become specialists were appointed and then supported under EQRPI through training and ‘learning on the job’. In particular, leading test developers emerged from the project’s subject working groups, which had been formed from the best subject specialists from respected schools, universities, and the NIE. However, it should be noted that recruiting and retaining the best technical staff in the ATC remains a challenge.
Since the ATC has been fully operational, it has become a key mechanism for enhancing the scope and quality of the overall assessment system in Armenia. This is largely due to two key factors. First, as the organization has proved its competence, the MOES has found it increasingly convenient to transfer responsibility for examinations and other forms of assessment to ATC. Second, building on capacities developed through consultancy and training services supplied under EQRPI, the professional staff of ATC are now initiating qualitative reforms. In particular, there are signs that ATC is strengthening its analysis, reporting, and feedback mechanisms.

**Drivers of Change**

In Armenia, as in many other former socialist republics, reforms in the field of examinations were given impetus by two powerful internal drivers. First, GORA and the MOES recognized the importance of reforming high-stake examinations in order to reinforce broader efforts being made to improve the content and delivery of the national curriculum to make it more relevant to a modern, global, competitive environment. Second, there was great concern that traditional university admissions procedures were not only inequitable and opaque, but also prone to corruption and malpractice. In short, there was a clear understanding of the role that such examinations play both within the educational environment and in society as a whole. This was in tune with the view of international agencies that external examinations provide a powerful lever for bringing about desirable changes in an education system. This may be summarized as: “Good examination design can lead to social cohesion, it can help choose talent fairly, it can engender trust in public institutions, and it can give confidence in the general conduct of the public school system” (Heyneman 2009, p.11). The alignment of internal and external drivers ensured that the design of EQRPI allocated sufficient human, physical, and financial resources for building a new examination system and, of supreme importance, a new organization to deliver that system.

Inevitably, a tension exists between designing examinations to enhance educational quality and designing them to fight corruption. In Armenia, the most important driver in the introduction of the UE was the need to maintain security and eliminate, as far as possible, malpractice in
the assessment of students and hence, the university admissions system. Therefore, in common with many other transitional countries in the region, all UE tests consist entirely of objective, selection-type items whereby students’ selected responses can be recognized by optical character recognition software and scored by computer systems. The limitations of this approach, and the likely negative impact on teaching, were well understood by those who designed the system. However, in the early years, it was essential that the integrity of the system should be maintained and so the examinations were introduced in this ‘safe mode’. Now that the system has become embedded and gained a degree of public trust, some commentators have suggested that changes should be introduced to increase the educational validity of the examination, but the tension between maintaining security and enhancing educational value is still strong. This is captured in the following observation:

There is … a serious risk that the educational value of the exam will not be considered the main priority, and question types which can promote student learning may be neglected. These include questions or tasks requiring oral fluency, skills in solving non-routine problems and problems related to “real-world” situations, writing and practical work. Many of these latter skills are regarded as essential for preparing students for the knowledge economy and for the development of important “soft skills.” The priority given to security could lead to an examination system dominated by the exclusive use of computer-scored multiple-choice questions (READ 2012, p. 7-8).

Unfortunately, from an educational viewpoint, it appears that the obsession with maintaining security and eliminating all subjectivity may be winning the argument. For example, as described previously, all objective items included in UE question papers are now selected from an open item bank, effectively eliminating the opportunity to test higher-level skills in novel, authentic contexts.
Large-Scale Assessments

International Large-Scale Assessments

Armenia participated in the 2003, 2007, and 2011 cycles of TIMSS, with samples being drawn for both the Grade 4 and Grade 8 populations. It is also participating in the 2015 cycle. The main purposes of Armenia’s participation in TIMSS are: (a) to give a snapshot of what Armenian students know and can do in the subject areas of Mathematics and Science at Grades 4 and 8; (b) to provide data that can be used to identify trends in educational performance, i.e., to track progress over time; (c) to allow the achievement levels of Armenian students to be compared with those of their peers in other countries; and (d) to identify factors that may promote or hinder educational progress so that these can be accounted for in national policies, strategies, and practices.

In TIMSS, as in all major international, large-scale assessments (ILSA), the highly technical process of drawing a representative sample is carried out by an international contractor, using the sampling frame provided by the country in question. A typical sample for Armenia at each population assessed is 150 schools and approximately 5,500 students. For example, Armenia’s achieved sample in TIMSS 2011 was 5,146 students in 150 schools for Grade 4, and 5,846 students in 153 schools for Grade 8 (IEA, 2012).

The tests are developed by the international contractor according to the TIMSS assessment frameworks and procedures. Items are translated and locally field tested. The degree to which the TIMSS assessment frameworks are aligned with national curricula varies across countries. In Armenia, the correspondence in Mathematics is relatively good, especially at the Grade 8 level where all 19 topics in the 2011 TIMSS framework were explicitly required by the national curriculum. Alignment is less close for Science, but is still considered adequate.

---

21 Armenia also took part in the TIMSS Advanced study of 2008, which is targeted at students studying Physics and Mathematics in the final year of compulsory general education.

22 Assessment frameworks are available at: http://timssandpirls.bc.edu/timss2015/frameworks.html

George Bethell and Karine Harutyunyan
The data for Armenia presented in the international reports of the International Association for the Evaluation of Educational Achievement (IEA) are the main source of information used to report the nation’s performance to policy makers, educational practitioners, and the wider public.\textsuperscript{23} For example, data extracted from the international report is used to disseminate headline results via the media and through conferences and seminars aimed at educational professionals.

A national report on Armenia’s performance in TIMSS is produced soon after completion of the test administration and scoring phase, and before the international report is released.\textsuperscript{24} The report includes national distributions of student achievement in Mathematics and Sciences as raw scores, and a comparison of the results by administrative region (marz); correlations between achievement and background factors; and comparison with previous TIMSS results prior to scaling, i.e., raw scores.\textsuperscript{25} Since all preliminary analyses are based on raw scores, no comparison of Armenia’s results with those of other countries is included. This report is not widely circulated and it is unclear whether it has much of an impact on policy makers or practitioners.

**National Large-Scale Assessments**

Building on experience gathered through TIMSS, Armenia is now implementing a NLSA program across three domains: Armenian Studies (Armenian Language and Literature, and Armenian History), Science (Physics/Chemistry and Geography/Biology), and Foreign Languages (English and Russian). Four surveys have been completed to date: Armenian Studies (HAAS) in 2010, Physics and Chemistry (BAAS) in 2011, Geography and Biology (BAAS) in 2012, and Foreign Languages (OLAAS) in 2013.\textsuperscript{26} NLSA studies differ considerably from all other assessments previously conducted in Armenia in their purposes, design,
and instruments. In particular, the items are not limited to assessing knowledge from the curriculum, but also assess the practical application of knowledge and skills relevant to everyday life. The HAAS and BAAS tests include multiple-choice, short-answer, and open-ended questions whilst the OLAAS tests include text-based items similar to those used in, for example, PIRLS.

The tools for NLSA follow the general model used in TIMSS in that each student selected to take the test completes three parts: a test for the first subject, a test for the second subject, and a questionnaire designed to collect information on student background. Information about NLSA instruments is available on the ATC website, including some examples of the tasks presented to students.

Typically, six test booklets linked by common (anchor) items are used for each NLSA. Stratified samples are selected, but not by probabilistic methods. Analysis of results focuses on raw scores (e.g., percentage correct), but item response theory (IRT) is used to ascertain the psychometric properties of items. Technical reports are produced, but these are predominantly for ATC’s internal use. They are available to MOES staff, but are not yet used to inform teachers and other practitioners. It should be noted that ATC considers its work thus far on NLSA to be “research and development—just the first step towards an effective NLSA” (Bagdasaryan, personal communication, 2014).

In addition to the national assessments described above, another form of sample-based assessment—so-called ‘External Evaluations’—has been carried out by ATC since 2010. The purpose of these evaluations is to provide objective data about student achievement in different subject areas that can be compared with the results from more subjective, internal school assessments. The evaluations cover Grades 5-11, with subjects and grades to be tested changing annually. The testing instruments are closely based on the curriculum, and are very similar to the tests used for the state graduation examinations in their structure, methods, and scoring procedures. However, they are applied to a sample of students from a limited number of schools. Pragmatic, non-probability sampling methods are used, but the sample is considered to be adequately representative of the population in terms of school type, size, and location. In addition to tests, questionnaires are completed by school
directors and teachers in order to investigate the relationship between student achievement and certain background factors.

In 2013, the External Evaluation was conducted in 63 schools selected proportionally from all marzes (administrative regions). Approximately 2,700 students participated in tests of Armenian Language and Literature, and Mathematics. 267 teachers were also included in the survey. Each subject test lasted 60 minutes and comprised 20 dichotomously-scored items. The scoring scale was 0-20 points, with final results transformed to the ubiquitous 10-point reporting scale.

The results of External Evaluations are analyzed and reports are produced. The 2013 report is available on the ATC website.\textsuperscript{27} The findings of the External Evaluation are usually discussed at the Ministry Collegium. It is difficult to know what, if any, impact the information from this form of assessment has on the wider system.

**Current Policies, Structures, and Resources for ILSA and NLSA**

Government policy regarding ILSA indicates only that Armenia will participate in the 2015 cycle of TIMSS. In recent years, it has been proposed that Armenia should also take part in the Progress in International Reading Literacy Study (PIRLS) 2016 and the Organization for Economic Cooperation and Development (OECD) Program for International Student Assessment (PISA). However, these have not been approved by the GORA and so the de facto policy is that, for the foreseeable future, Armenia’s participation in ILSA will be limited to TIMSS.

ATC is the institution responsible for the design, conduct, analysis, and reporting of TIMSS and all forms of national assessment. It has a staff of 81, including 15 assessment specialists, who work on both NLSA and ILSA activities according to need. Its structure is organized by staff functions and expertise rather than according to the type of assessment and so it is difficult to estimate with precision the human resources associated with specific large-scale assessments.\textsuperscript{28}

\textsuperscript{27} Available only in Armenian.

\textsuperscript{28} This is not the case in, for example, Lithuania where a dedicated unit for ILSA sits within the National Examinations Centre.
ATC has to fund all NLSA activities from its total budgetary allocation for contractual works (i.e., production and processing of examinations and large-scale assessments). In 2013, the provision for all such works was AMD 167 million (~USD 400,000). It is difficult to estimate exactly how much of that was allocated to a particular NLSA.

For a relatively poor country such as Armenia, participation in TIMSS is expensive. In addition to the standard international fees required for participation in the exercise, there are significant national costs associated with translation, printing, packing and distribution, scoring, and data entry. For example, Armenia’s national budget for the 2011 TIMSS cycle was AMD 15,450,000, equivalent to about USD 40,500. Approximately one-third of this was used for field testing and two-thirds for the main data collection. In addition to these costs, there are significant hidden costs associated with services provided ‘free of charge’ by, for example, regional educational offices, schools, principals, and teachers.

**Mechanisms of Change**

The main vehicle for establishing the foundations of Armenia’s comprehensive student assessment system—including ILSA—was EQRP I, which was largely funded from a World Bank credit agreed in 2003. EQRP I provided resources for building operational capacity and establishing the ATC, which eventually would be available to carry out system monitoring assessments. It also provided essential funding for Armenia’s participation in the 2003 cycle of TIMSS.

Prior to ATC becoming a fully-functioning organization, the CFEP was the mechanism through which the 2003 and 2007 cycles of TIMSS were conducted. This was a very unusual arrangement given that the implementation agency is usually a technical institution (e.g., a university department or a national assessment agency). In any event, the CFEP collaborated effectively with the National Research Coordinator (NRC) and was successful in commissioning all services necessary to prepare materials and conduct testing activities in the field. This transitional arrangement proved highly effective and once ATC was operational, on-going responsibility for TIMSS was transferred to it.

The mechanism for funding ILSA has changed over time. Initially, Armenia was exempted by the IEA from the payment of the international
costs associated with participation in TIMSS, with World Bank project funds covering the remaining national costs for TIMSS 2003 and 2007. For the first half of the TIMSS 2011 cycle, international costs were paid from EQRP I funds. Subsequently, national costs for the field testing and main data collection activities and international costs for the second half of the cycle were provided by the State through the ATC annual budget. Once again, the transitional arrangements worked well and now all ILSA costs are met through ATC’s operational budget.

**Drivers of Change**

Initially, the fundamental driver of change was GORA’s commitment to the wholesale restructuring and reform of the legacy Soviet system to ensure that national educational outcomes would be appropriate for a modern country competing in a global market. It was recognized that this would require an international perspective and comparative data on educational standards and trends. An external driver that resonated with GORA’s objectives was provided by the World Bank which, at the time of planning for EQRP I, was increasingly advocating that countries investing in education should participate in ILSA in order to gain evidence to inform policy making. TIMSS was, at the time, the predominant ILSA and being curriculum-based was the obvious choice for Armenia.

The development of capacity within ATC under EQRP I, and through participation in TIMSS, contributed to the establishment of a professional assessment community where, to all intents and purposes, none existed before. Similarly, the involvement of the MOES in the project generated increased interest in the use of assessment-related data for system monitoring. The combination of the two served as an internal driver for the development of the two forms of NLSA currently being implemented by the ATC.
Discussion and Lessons Learned

Summary of Progress Made
Over a period of little more than two decades, Armenia has made great progress in its efforts to develop a comprehensive system of student assessment to monitor and enhance educational outcomes across the general phase of education. Advances have been made on all fronts: the university admissions process has been made fairer and more transparent through the introduction of the UE; teaching and learning have been enhanced through the introduction of better classroom assessment practices; and national and international sample-based assessments are starting to provide potentially valuable information to educational policy makers and practitioners. The ATC now serves as a sustainable national center of excellence for assessment activities, and around it, a professional assessment community is emerging with the potential to maintain and further strengthen assessment practices. Much remains to be done in all of these areas, but the foundations are firm.

Key Drivers
The drivers behind Armenia’s progress in this field have been both external and internal. Externally, agencies such as the World Bank have consistently reinforced the message, based on international evidence, that for a modern education system to be effective and efficient, a well-designed student assessment system is needed to complement all of the other necessary elements of the teaching/learning process. This message was well received by GORA, which proposed that credit agreements for educational reform projects should include significant funds explicitly for the enhancement of educational assessment. In Armenia, projects—most notably EQR P I—have provided resources and a structure for driving forward assessment-related reforms. However, external drivers on their own are never sufficient. Indeed, it is undoubtedly the internal drivers that ultimately determine the direction of travel and the rate of progress. In Armenia, the most powerful driver was and remains the government’s commitment to improving education in order to strengthen the country’s most valuable economic resource—its people. This can be seen in its willingness to invest heavily
not just in infrastructure, but also in the ‘softer’ areas of education, such as the development of modern subject curricula, learning standards, and assessment practices.

A secondary internal driver, but one with a very high profile, was the government’s commitment to reducing the perceived level of corruption in assessment practices. This has been an objective shared by many former socialist republics, and the first step is almost always to reform traditional university admissions procedures through the introduction of a secure, centralized examination system. Inevitably, this has both positive and negative consequences, although the immediate benefits for society (e.g., greater equity in access to HEI) usually outweigh the disadvantages. In Armenia, the drive to reduce corruption through the development of the UE led indirectly to the establishment of the ATC—a key mechanism in the reform of general assessment practices.

A third internal driver was the Government’s desire to align assessment practices more closely with international best practice. In this, it saw the potential to increase fairness, transparency, and objectivity in decision making through the use of modern technologies. It also recognized the need to develop a cadre of assessment professionals in order to implement and sustain the desired reforms.

In the case of Armenia, clearly-articulated goals and objectives resonated strongly with external drivers, giving greater impetus to reform initiatives and increasing the chances of success.

**Key Mechanisms for Bringing About Reform**

The most effective mechanism for bringing about reform in the area of student assessment has undoubtedly been the GORA’s judicious use of educational projects to implement initiatives in a range of well-defined areas: high-stake examinations, classroom assessment, teacher training in assessment, and participation in international studies (TIMSS). Funds made available through credit agreements enabled GORA to support capital expenditure in, for example, establishing the ATC as a fully-operational organization at a time when the national budget was under great pressure from many competing interests. Subsequently, operational costs associated with educational assessment have been transferred to appropriate state budget lines, ensuring long-term sustainability.
Maximizing the effectiveness of such projects requires careful planning prior to implementation and good management during implementation. The mechanism for ensuring that both aspects were adequately addressed was through the creation of a designated project implementation unit (i.e., the CFEP) in the MOES. The responsibilities of the CFEP included not only the administrative and financial elements of project management, but also monitoring the quality of professional services and activities and their alignment with MOES expectations.

International consultants were, and continue to be, a major mechanism in implementing assessment reforms in Armenia. Some countries are reluctant to use international expertise because of the difficulty of ensuring value for money. In Armenia, there is a willingness to use international technical assistance provided it is of high quality and provided that local control can be maintained.

Another prominent mechanism in transferring knowledge and skills has been the significant investment in training across a number of assessment-related areas and for a wide range of beneficiaries. Training has been provided both with international technical assistance and through local providers such as the NIE. Under EQRPI, one strand was to build professional capacity through training ATC staff and associate test developers (mainly from universities) in the principles and practices of assessment and measurement. A second, complementary strand was the provision of training in classroom assessment techniques. Here, lead trainers prepared under the project provided local, school-based training which reached, to all intents and purposes, every teacher. This commitment to training continues with the further strengthening of in-service professional development for teachers and the preparation of university staff to deliver the pre-service module on assessment.

Finally, the ATC established under EQRPI now serves as a major mechanism in implementing and sustaining reforms in the assessment system. GORA has invested heavily in ATC’s accommodation, equipment, and staff development. It continues to fund the organization from a dedicated line in the state budget. In return, ATC now offers a natural home for all the examination and assessment activities currently required by the MOES and it has the capacity to take on new initiatives in the future. The effectiveness of this has already been shown through the gradual transfer of additional responsibilities to ATC after the successful
introduction of the UE, including responsibility for international and national large-scale surveys of learner achievement.

**Selected Lessons Learned and Associated Challenges**

*Lesson 1: The technical elements of new forms of assessment are relatively easy to introduce; it is far more difficult to ensure that the information they generate is used effectively by policy makers and practitioners.*

At the start of Armenia’s educational reforms, the capacity to design and implement high-quality examinations and sample-based surveys of learner achievement was extremely limited. However, through the mechanisms described in this paper, the country has, amongst other things, successfully introduced a high-stakes examination for university selection and conducted three cycles of an ILSA. Armenia is not unique in this. Since 1991, practically all former Soviet republics have established centralized examination systems from scratch (Bethell and Zabulionis, 2012). The test papers may not always be of the highest quality, and one may question the educational validity of some of the procedures used, but from a technical point of view the new systems are operational and, in general, have few major problems. In addition, many former Soviet republics have participated in at least one major ILSA and, with a few exceptions, have successfully met international procedural standards.

The conclusion is clear: with sufficient investment and preparation, countries do not find it difficult to build the technical capacity to ‘do’ an assessment. However, examinations and other forms of assessment are conducted in order to serve specific purposes, *including the provision of potentially valuable information for informing the decisions of policy makers and guiding teachers and other practitioners.* It is this element that is often neglected or badly implemented.

In Armenia, the state invests a significant sum from its budget in order to participate in TIMSS. The country’s results are presented in the international report, allowing some basic comparisons to be made. There is also a preliminary national report that summarizes student performance based on raw scores, but it is not clear what value this has. No further secondary analysis is conducted, no independent research is commissioned, and no subject-specific reports are prepared for schools and teachers. As a consequence, much potentially valuable information is left unused. This is not unusual, as evidenced by the following statement
from an OECD Education Working Paper: “in most low- and middle-income countries, there is no PISA national report and no national data analysis. Reasons vary, from low analytical capacity to human and financial resource restrictions, or lack of political interest.” (Bloem, 2013, p.22).

The OECD statement reveals two critical dimensions in establishing an effective feedback system: political interest and analytical capacity. The first of these generates the demand for information and, in particular, relevant information presented in an appropriate form. The second supplies the information demanded in the required form. The simplest scenario is uni-directional, i.e., the MOES asks ATC to supply relevant information and ATC conducts the analysis to provide it. However, this will not work if policy makers do not know what to ask for, perhaps because they do not know what is available or how they can use such data. Therefore, in practice, the relationship should be bi-directional: assessment specialists within ATC should generate additional demand by providing policy makers with the richest set of data possible presented in user-friendly formats. Once policy makers see the potential benefits of assessment-related data, it is likely that they will ask for more, thereby reinforcing the feedback loop. Achieving this requires action on three fronts:

- educating policy makers so that they can interpret assessment-related data intelligently;
- motivating assessment specialists so that they generate more and better data; and,
- providing resources so that secondary analysis and focused research can be conducted.

In addition to providing data to policy makers, the assessment agency (ATC in the case of Armenia) should provide feedback to subject teachers so that they can better understand the strengths and weaknesses of typical learners and hence improve their teaching practices. For teachers, raw scores and other descriptive statistics presented out of context are of little, if any, value. Therefore, subject specialists with an understanding of assessment statistics need to select and present data in forms that teachers can use. For example, TIMSS international reports include a limited number of exemplar items for Mathematics and Science
with data on the proportion of students responding correctly in each participating country.

Examples of good practice in national sample-based assessments include the comprehensive subject reports produced from data gathered in the National Assessment of Educational Progress (NAEP) conducted in the USA. These include ‘item maps’ that set out, using concrete examples, which tasks have been mastered by students at different parts of the ability range.²⁹

In this feedback loop, the community of teachers is unlikely to initiate demand. Therefore, the assessment agency needs to be proactive and assume responsibility for providing useful information in appropriate formats. It also needs to contribute, perhaps through collaboration with other organizations, to the training of teachers in interpreting assessment-related data.

Lesson 2: In the case of examinations, where the stakes are high, there is inevitably a tension between the need to maintain examination security and the desire to promote important educational objectives through more innovative and effective approaches to examination design.

Armenia, in common with many other post-socialist states, introduced centralized examinations for university selection purposes in a climate where the perception of corruption is great.³⁰ The pressing need to make the new system objective, transparent, and secure dictated the use of so-called ‘technological’ items capable of being scored by computerized systems. The obsession with transparency also led to the adoption of an overly-simple scoring and reporting scale, as previously described. These measures, coupled with the use of highly visible, anti-cheating measures in testing centers ensured that Armenia’s UE was successfully introduced in ‘safe mode.’ However, placing the emphasis on anti-corruption inevitably relegated certain educational objectives, e.g., promoting the teaching and learning of higher-level cognitive skills, to second place. The presumption was that once the new system had embedded, more valid forms of assessment, such as the use of problem-solving tasks set in novel

²⁹ Subject reports for NAEP can be found at http://nces.ed.gov/nationsreportcard/ (accessed 13 July 2015).
contexts and extended-response items (written components) could be introduced. In reality, the concern with security and transparency has not receded. On the contrary, recent policy interventions—in particular the move towards the use of open item banks—have shifted the system further away from highly desirable educational values towards mechanistic transparency and anti-corruption measures. There are undoubtedly some short-term gains with, for example, fewer complaints being lodged about the examination questions. However, what will be the long-term impact on national educational outcomes? Will the young men and women who emerge from an education system dominated by rote learning have the higher-level skills required by a modern workplace in an increasingly competitive global market?

This tension between maintaining security and harnessing the potential positive backwash effect of high-stake examinations will persist while society’s perception of corruption remains high. There are no simple solutions. However, policy makers should be aware that an appropriate balance is required if desired educational outcomes are to be promoted. Furthermore, it is the role of the assessment community to convey this message and to propose suitable strategies.

Lesson 3: Assessments are not isolated from other aspects of society. As the context in which they operate changes, there will be a need to review their ‘fitness for purpose’ and, where necessary, adapt them accordingly.

When evaluating student assessment systems, it is all too easy to get caught up in the technical aspects and to forget that assessment is not an end in its own right—it is merely a tool that should serve the education system and society as a whole. In this subservient role, assessment systems should be dynamic and respond to changing socio-economic contexts and shifting priorities. Armenia’s UE provides an interesting example of this.

University selection examinations are a mechanism by which demand for further education is matched with supply. When the UE was introduced in 2007/2008, competition for places in popular courses in HEI was significant, especially for state-funded places. Where university faculties faced the problem of numerus clausus, scores from UE subject tests provided objective information for allocating both places and any available scholarships. Since then, however, the nation’s demography has
changed significantly due to a number of factors, including falling birth rates and increasing migration. For example, total enrolment for the general phase of education in state and private schools has fallen from 411,439 in 2010/2011 to 359,966 in 2013/2014—a drop of 13 percent (National Center of Educational Technologies [NCET], 2015). The change is even more startling when enrolment in the year of graduation is considered with, as shown in Figure 2, a drop of more than 36 percent since 2010/2011.\textsuperscript{31}

**Figure 2: Annual enrolment data for the final year of general secondary education (G11 in 2010/2011 and G12 thereafter).**

![Figure 2: Annual enrolment data for the final year of general secondary education (G11 in 2010/2011 and G12 thereafter).](image)

*Source: National Center for Educational Technologies at http://stat.armedu.am/*.

With the school population falling rapidly, the absolute demand for university places fell, but the number of places was not reduced accordingly. With relative supply increasing, competition was less fierce and the perceived need for the selection examination was reduced. To complicate the matter further, this demographic shift took place against a

\textsuperscript{31} In the year 2010/2011, students graduated in Grade 11. Since then students have graduated in Grade 12.
backdrop of declining economic strength, which caused the government to reduce funding for HEI by cutting the number of scholarships, especially for popular specialties such as Economics and Law where demand was still great. Universities responded, not by cutting costs by closing under-subscribed courses, but by increasing the number of places for fee-paying students on popular courses in order to maintain income. This increased the supply in absolute terms which, in turn, reduced competition still further.

From 2007 to the present, HEI enrolment rates in Armenia have increased. This may be a desirable outcome for a country with a very large cohort, but in Armenia, admitting a larger proportion of students means that the average quality of the intake to HEI is lower than in previous years. The impact is most noticeable in less popular, but nationally vital, courses such as Agriculture, Science (Biology and Chemistry), and Engineering where students are now regularly admitted with minimum passing scores of 7.5 out of 20 in one subject only.\footnote{This is equivalent to a score of just 30 out of 80 on tests having significant ‘guessing scores’ since they are composed mainly of four-option multiple-choice items. In Geography, for example, the ‘blind guessing score’ from the multiple-choice items is 19/80.} In this situation, UE does not serve as an effective filter either on numbers or quality. This has led at least one leading Armenian educator to question whether any form of selection by merit is needed at the school/university interface!

The lesson here is that the relationships between supply, demand, and quality in higher education are complex and so policy makers need to understand fully the consequences of any decisions they make with regards to (a) the nature of the selection instruments and (b) the minimum criteria for success.

**Conclusion**

Through its sustained efforts and prudent investment, Armenia has established, in little more than two decades, the foundations of a comprehensive and coherent system of student assessment. Along the way it has also developed the core of a potentially vibrant assessment
community. However, many challenges remain in maximizing the effectiveness of the system and ensuring it best serves society. Meeting these challenges will require the engagement and action not only of the ATC, but also policy makers and the broader community of educational practitioners. None of these challenges is unique to Armenia. Armenia can learn from the experiences of others, and others can learn from the experiences of Armenia.

References


List of papers published in the current Working Paper series


Forthcoming papers

In order to maximize effectiveness, every national education system needs a comprehensive and coherent approach to student assessment. The SABER framework for student assessment identifies four key elements in this regard: classroom assessment, examinations, national large-scale assessment, and international large-scale assessment. Armenia has undoubtedly made significant progress in all of these areas since gaining independence from the Soviet Union in 1991. The purpose of this paper is to describe Armenia’s experiences, share the lessons learned along the way, and point out the challenges that countries face in trying to ensure that their assessment systems remain fit for purpose in a changing environment.

George Bethell, Independent consultant specializing in the policy and practices of educational assessment

Karine Harutyunyan, Deputy Minister in the Armenian Ministry of Education and Science