



Year 2012

Project:

Development of Final Examinations at the End of Educational Cycles in Croatia

Summary Report to
National Centre for External Evaluation of Education
1 January 2012 – 31 December 2012



**ASSESSMENT GROUP
INTERNATIONAL, SPRL**



Eduardo C. Cascallar, Ph.D.
Managing Director

SUMMARY REPORT OF ACTIVITIES AND OUTCOMES 2012

Project:
**Development of Final Examinations at the
End of Educational Cycles in Croatia**



ACKNOWLEDGMENTS

During this second year of the project Assessment Group International conducted a series of activities in support of the Development of Final Examinations at the End of Educational Cycles in Croatia project, in conjunction with, and with the support of the National Centre for External Evaluation of Education of Croatia. Special thanks are due to all the staff and authorities at the Center, which implemented much of the work, and which achieved very positive results in all tasks corresponding to this year's work.

All the many accomplishments of this year's activities would not have been possible without the expert work and support of all the Centre's staff and authorities to the activities conducted by Assessment Group International, in particular we wish to acknowledge the excellent work of the test coordinators and psychometric staff, as well as of the many teachers and local experts without whom our work would not have been possible. Special thanks are also due to the Director, Mr. Goran Sirovatka, to the Assistant Director, Ms. Jasmina Muraja, and to the Project Manager, Ms. Natalija Gjeri, for their unwavering support and facilitation at every stage of the project.



ACTIVITIES CARRIED OUT DURING 2012

1. PREPARATIONS AND IMPLEMENTATION OF THE ITEM WRITING WORKSHOP

a) Meetings to finalize the analysis of item writing needs given the examination architecture and blueprint for each of the four FEPE examinations

This activity involved separate meetings with each of the four content area coordinators and local psychometric staff, corresponding to each of the final examinations, to determine item writing requirements and specifications, taking into account the architecture and the blueprint of each FEPE.

The outcome of this task was to finalize an item writing requisition statement, corresponding to the structure and weights determined for each of the FEPE, which had specific requirements for the type and number of items needed for each examination, taking into account content area, and subarea, as well as expected difficulty and cognitive level.

b) Item Writing Workshop

This activity consisted of the organization of a workshop for content area expert groups, psychometricians and other center staff.

The outcome resulted in the training of the participants on item writing techniques to achieve a better distribution of items in all levels of cognitive processing, and difficulty levels, to improve the quality of the multiple choice examinations. In addition, issues corresponding to proper development of essay items and short responses were addressed.

c) Consultation on progress of item review status of items received

The core activity during this period, entirely conducted by NCEEE, was to achieve an adequate review of the items received from the commissioned item writers. With this goal, and in order to guide the review of the items received, several separate working sessions with the participation of each content area group were organized by NCEEE. In addition, a Skype session was organized between the Project Manager and Assessment Group International, to follow up on issues arising from this review process and update on status of items received.

These activities were very successful and established an initial item pool for each of the four FEPE after a comprehensive review of the items received.

2. IRT TRAINING AND OTHER CONSULTATIONS

One of the major goals of the project is to train local staff, building local capacity, to create a cadre of experts and other staff that could eventually run this and other similar testing programmes. With this objective in mind, a series of lectures and consultancy meetings were organized at NCEEE on psychometric underpinnings including applied IRT topics, software packages, and issues in test design for the correct development of the architecture of the exams. Issues in examinations equating and data analysis techniques were also included in the training agenda in a different workshop.

The outcome of these lectures and consultation meetings centered on IRT methods and the Parscale application software, was an excellent beginning in the direction of building local capacity, with the attainment of the expected general level of training in selected NCEEE test development and psychometric staff and other selected professionals.

Some of the topics that were discussed and covered during the consultation meetings and lectures of the year included:

- o FEPE examination design**
 - Planned architecture of the tests
 - Conceptualization of the exams with input from the various stakeholders

- o Development of the FEPE examinations:**
 - Item development
 - Item development workshops
 - Item review
 - Pilot and Field testing of items
 - Development of sufficient number of pre-equated test forms in each of the content areas.
 - Design and construction of pre-equated forms.
 - Design and implementation of field tests
 - Review of pre-equating of test forms
 - Development and distribution of testing booklets

- o Marking**
 - Scanning and scoring of multiple-choice items

- Cleaning and preparation of data
 - Development of scoring scales for open-ended items
 - Training of markers for open-ended items
 - Procedures for marking
 - Quality control of marking process
- **Introduction to the concept of standard settings**
 - **Standard setting sessions for each of the subjects tested in the FEPE examinations**
 - Selection of judges
 - Planning and carrying out of standard setting sessions
 - Training of assistants to coordinator of standard setting sessions
 - Use of corresponding modified-Angoff methods for multiple-choice and open-ended items
 - Determination and recommendation of achievement levels for the examinations
 - **Logistics and other procedures**
 - Examinations steps, processes and logistics
 - Test administration and registration
 - Logistics for marking of open-ended items
 - Data archiving; relationship to item banking operations
 - Appeals (registrations of requests for appeals, re-marking of tests, data entry of new tests results, and publication of results).
 - **Security procedures**
 - Safeguarding of all items and test forms.
 - **Comparability of tests across years: equating designs, vertical scaling planning, anchor design, etc.**
 - **Psychometric analysis: IRT based implementation of various aspects of the data analysis, test development, scoring, and test reporting.**
 - Development of a test design;
 - Classical test analysis of test data;
 - Extension of classical test theory analysis;
 - Introduction of Parscale software;
 - Need for Item Response Theory;

- Introduction of Item Response Theory;
 - Introduction of IRT analysis using 1P- 2P- and 3P models.
 - Correct standard error calculations for test that could combine open-ended and multiple-choice item types
 - Use of IRT methodology for the analysis of field testing and pilot testing results
 - Advantages of form-builder software to construct.
- **Topics regarding test reporting and test and score use**
- Taking into consideration all stakeholders involved: NCEEE, students, parents, teachers, MoE, and other government structures concerned. Reporting of examinations results, addressing test and score use; adaptation of the reports for the various stakeholders.

3. LECTURES ON DEVELOPMENT OF STANDARDS AND TRAINING IN THE DEVELOPMENT OF PERFORMANCE LEVEL DESCRIPTORS FOR THE FEPE DEVELOPED FOR 4 SUBJECT AREAS

Lectures on the development of standards

The central objective of this activity was to build capacity at NCEEE among content experts and psychometric staff through lectures in the development of standards.

With this goal in mind, lectures were organized at NCEEE on the development of content and performance standards.

This activity was very successful, with the attainment of the expected general level of knowledge in selected NCEEE staff, teachers who are members of test development teams, and other selected professionals.

a) Lectures on the development of Content Standards

In this activity the focus was on how the development of content standards is carried out. Content defines the various areas of knowledge and skills which all the students should learn, and content standards define those parts of the content in the curriculum that all the students should master. Although it was not the task of this project to develop these standards, it was important to instruct on how they are developed and their purpose.

b) Lectures on the development of Performance Standards and training in the development of level descriptors

Performance standards, on the other hand, describe how well the students should perform in these content areas. It is important to work closely together with key stakeholders involved in order to define and set performance standards. This training explained the underpinnings of such activity, and a separate workshop provided training on the process for the development of performance level descriptors as it was going to be implemented in the FEPE programme.

These two activities offered the opportunity to cover issues of good practice, innovations, and theoretical and practical procedures in the development of content and performance standards.

In both of the activities described above, results were highly successful, with the attainment of the expected general level of training in selected NCEEE staff, teachers who are members of test development teams, and other selected professionals, in the development of performance level descriptors

4. IRT EQUATING AND VERTICAL SCALING WORKSHOP

This activity involved the conducting of a workshop that covered various technical and practical aspects of IRT equating procedures and techniques, and the application of those techniques for the purpose of vertical scaling. Both theoretical and practical considerations were covered, as well as topics that relate to the planning and design of a testing programme in order to make equating possible, and the design of an assessment plan (across grades) to eventually make vertical scaling a possibility. Procedures and different types of equating approaches were covered, as well as various ways of using and reporting results, both for accountability at the level of central educational authorities and as formative feedback to schools and other institutions.

5. FINAL REVIEW OF ITEMS AND CONSULTATION ON SPECIFICATIONS/STATISTICS OF ITEMS RECEIVED

a) Review of commissioned items received – Review and construction of initial item pool.

The objective for this activity was to finalize the review of items that had been commissioned and received. This was carried out by NCEEE staff, test coordinators and psychometric staff, together with teachers who were members of the test development groups for each of the

FEPE examinations. Assessment Group International had provided consultation regarding item characteristics and quality indicators, relevant to this activity during the course of the project. These reviews, carried out entirely by NCEEE, were performed in separate sessions with the participation of each content area group and one local psychometrician per group.

As expected, this activity resulted in the final construction of the initial item pool for the four FEPE that was to be used in the pilot administration. In addition, a final check of item correspondence with content and estimated psychometric design parameters was completed. These efforts resulted in a final item pool for the initial administration of the four FEPE.

b) Consultation regarding psychometric training at NCEEE

The goal of this consultation was to better design a training programme of the NCEEE psychometric staff, in order to maximize the capacity building at the Center, both for the purposes of this particular project and for further test development, data analysis, and implementation of programmes.

The topics considered involved:

- Determination of needs and requirements
- Planning of training
- Form construction
- Software training with Parscale and specific IRT topics (chain- and concurrent-calibration, equating, handling of item parameters, etc.)

It resulted in the planning of an advanced workshop that followed:

- Advanced Psychometric workshop for training in the use of Parscale software.

In addition, during the course of the project, AGI provided consultation at various meetings and lectures and training opportunities on principles guiding automated test construction, conceptual framework for an integrated engine for form construction, analysis, scoring and reporting.

c) Consulting on the development of forms/booklets, design and implications for analysis.

This consultancy was implemented to cover form/booklet design topics for NCEEE psychometric staff.

d) Advanced Psychometric Workshop

There was a workshop implemented to cover advanced topics in data analysis using the Parscale software. This workshop touched upon issues related to form construction and implications for analysis, item calibration, concurrent calibration, chained-calibration, anchor item construction, control for anchor item quality, evaluation of equating results, item evaluation from the IRT analysis, test evaluation from IRT analysis, possible ways of reporting IRT results. Both of these workshops included all of NCEEE psychometric staff.

The workshop was very successfully implemented and carried out. Although AGI has consulted NCEEE staff since the inception of this project, this training specifically focused on various practical advanced aspects of data analysis for the new FEPE examinations, including software (Parscale) utilization.

This Workshop involved the NCEEE psychometric staff and links were made between the various aspects of the data analysis process with topics regarding item characteristics, test quality and reporting needs. All trainings were coordinated with NCEEE administration and project management.

6. STANDARD SETTING SESSIONS (SS) IN ORDER TO ESTABLISH INITIAL ESTIMATES OF ITEM DIFFICULTY FOR THE ITEM POOL AND TO IMPLEMENT THE PILOT STUDY FORM CONSTRUCTION

a) Planning of the SS sessions

This necessary step was carried out both by NCEEE and AGI. The Center identified and selected the judges for each FEPE through a broad call to teachers from all over Croatia. Simultaneously, AGI prepared and planned the implementation of the SS, while NCEEE also carried out the logistical planning and preparing the materials to be used during the standard setting sessions.

b) Training of participants

At the beginning of each standard setting session, the essential aspects of the procedure were explained to the participants, as well as providing them with a brief overview of the theory and assumptions of the method. This was done in order to make them aware of important elements of the procedure chosen for the SS (a modified-Angoff method) and thus maximize the positive outcomes of the process.

c) Implementation of the SS sessions

Two AGI expert staff members carried out the standard setting sessions simultaneously, two SS-groups at a time, one for each examination, over a period of two weeks (three or four days per examination team for each SS, plus other days for preparation of materials).

d) Determination of initial cut-scores for each of the new FEPE examinations

A modified-Angoff approach was used for this stage of the project, in order to obtain provisional cut-scores for each one of the examinations, through a standard-setting methodology. It involved taking the difficulty estimates provided by each of the expert judges in each standard-setting panel (obtained through the informed partial consensus of the method), and estimating from them the suggested cut-scores for the initial pilot study. These cut-scores will be moderated by the actual results and the use of IRT methods to adjust the estimates once the pilot study is conducted and real data exists. In addition, it was recommended to use the Beuk adjustment technique, assuming a normal distribution, as an initial way of determining such provisional cut-scores.

Preliminary cut-scores were obtained from the standard-setting sessions, including the raw percentage of the test needed for entry in each of the corresponding performance level categories, as well as the expected pass rate (with the assumptions of normality used for the estimates). In addition, for each examination, the value of the adjusted percentage cut-score derived from the Beuk adjustment technique was also calculated and reported. These cut-scores will be moderated following the finalization of the pilot study, using the actual data obtained.

e) Planning and design of the Pilot study

This task has been going on from very early in the project. In this, the final stages of preparation, careful planning and logistical preparation took place, including the sampling of the school and student populations in the country according to various characteristics in order to guarantee a proper representation for the pilot study participation. All these activities were entirely carried out by NCEEE. Several other matters necessary for the implementation of the pilot study were carried out, with AGI support in the planning of the statistical analysis, including exercises of IRT-analysis simulating the upcoming examination structure and results. In addition, the determination of preliminary cut-scores was part of this process.

Overall Assessment of the TASKS for 2012

The level of professionalism and readiness at the Center has been very encouraging, and all tasks were completed satisfactorily and according to professional standards. The consulting role of AGI was encouraged and received the support of all staff and Center authorities. It has been extremely satisfying to observe the increase in local capacity at the Center, as well as the ability to implement a complex project in a very successful manner. Assessment Group International will continue to offer its expertise and know-how in support of the project, and is grateful for the opportunity to participate in such a valuable plan for the educational system in Croatia.

Development of final examinations at the end of educational cycles
Staff at NCEEE involved in the Examination Development

Name	Function
Natalija Gjeri	Project Manager
Zoran Žitnik	Project Coordinator
Content Area Coordinators	
Petra Chudy (starting May 2012); (Tea Vidović - until May 2012)	Content coordinators – Croatian language
Marijana Vučić, Vanda Petanjek, and Martina Prpić (from September 2012)	Content coordinators – English language
Mario Peranić, Irena Matoic, and Tajana Crnobrnja Miljković (from March 2012),	Content coordinators – History
Zlatko Zadelj, and Valentina Kapraljević (from March 2012)	Content coordinators – Physics
Psychometric Coordinators	
Ira Tretinjak	Psychometric coordinator – Croatian language
Tajana Lankaš (starting June 2012) (Natalija Ćurković - until June 2012)	Psychometric coordinator – English language
Josip Šabić	Psychometric coordinator – History
Zoran Žitnik	Psychometric coordinator – Physics