

# Large-scale Computer-based Testing of Foreign Language Competences across Europe

## Technical Requirements and Implementation

Jostein Ryssevik  
*ideas2evidence ltd, Norway*

### Abstract

*This article presents the software requirements of the European Survey on Language Competences (ESLC) – a large-scale international survey initiated by the European Commission to study the foreign language skills of European school children. The emerging open source software platform developed by the SurveyLang consortium is also described. The ESLC will be conducted in all or most EU member states and is aimed at pupils of lower secondary education studying one or both of the two most taught foreign languages in their country. The sampled schools will have the opportunity to choose between computer-based administration or a paper-and-pencil based equivalent. It is, however, a goal of the project to maximize the number of schools that choose the CBT-version and the software platform is designed to support this goal. The article describes the demanding technical requirements deriving from this dual-mode approach and the fact that the study is based on a complex incomplete block design where the delivered tests will be matched to the proficiency levels of the students. The software support for the item writing process is also described in this article, as well as the planned solution to the strict security requirements that always go hand in hand with international assessment surveys like ESLC.*

---

In order to develop a *European Indicator of Language Competence*, the European Commission has initiated a large-scale assessment survey to be conducted for the first time in the first quarter of 2011. The contract to develop, manage and analyse the output of the *European Survey on Language Competences*, was awarded to SurveyLang - a consortium of European partners headed by Cambridge ESOL and involving also among others the National Institute for Educational Measurement (CITO) and Gallup Europe. The development of the software platform is managed by Gallup Europe.

### ESLC – basic parameters

According to the European Commission's plans, the survey should be conducted in 32 countries across Europe, although it is still not decided whether all the invited countries will take part in the first round. The ESLC will test the competencies of European pupils in five languages - English, French, German, Spanish and Italian - but only the two most frequently taught languages out of these five will be tested in each country. The first round of the survey will focus on testing three language skills: listening, reading and writing. Measurements of speaking skills might be added in future rounds.

The target group of the study are pupils in the last year of lower secondary education or the second year of upper secondary education who are studying one of the two most taught foreign languages in their country. Due to major differences across countries when it comes to the introduction age of foreign language teaching, especially regarding the second foreign language, it is anticipated that the target group will have to include upper secondary pupils in some countries. In each country 1500 students for each of the two chosen languages will be randomly sampled to take the test. Each student will only be tested in one language and only in two of the three skills mentioned above.

In order to increase the precision of the tests and to avoid fatigue or boredom effects, it has been decided to introduce an element of targeting. As full-blown adaptive testing is hardly feasible and probably not desirable in a survey like the ESLC, a hybrid design has been developed based on a short routing test taken prior to the main survey. The routing test will classify the pupils in three proficiency levels. The information from the routing test will subsequently be used to allocate students across tests at different difficulty levels in a linked design that makes sure that each pupil receives at test that is targeted to his or her proficiency level. Combined with the principles of

an incomplete design where each pupil will only receive a portion of the test material that matches his or her proficiency level, targeted testing involves an extremely complex logistical scenario when it comes to the assembly and dissemination of unique testing sequences for each single student.

In order to reduce the burden on the participating countries, it has been decided to develop the ESLC for computer-based testing. However, given the variations in technological preconditions and computer skills across schools and countries, a paper-and-pencil based equivalent will also be offered. Besides the methodological challenges, this dual-mode design also increases the logistical and technical challenges. As will be described below, it necessitates a software platform that can support the administration of paper-and-pencil-based tests as well as of the computer-based tests in a coordinated way. An explicit objective is also to develop the software platform in such a way that as many schools as possible will be in a position to choose the computer-based alternative.

## Requirements

The technical and functional requirements of a software platform designed to support a large-scale survey operation like the ESLC are demanding. At a high level, the software platform should:

- support all the various stages and roles in the development and implementation of the survey (see Figure 1),
- enable the automation of error-prone and expensive manual processes,
- be flexible enough to handle the variety of test item types used by the survey,
- support the implementation of the complex incomplete block design described above,
- meet the high security requirements of international assessment surveys like the ESLC
- reduce the technical and administrative burden on the local administrators to a minimum
- run on existing hardware platforms in the schools
- be an open source platform available for free use by other actors (an explicit requirement in the terms of reference of this contract)

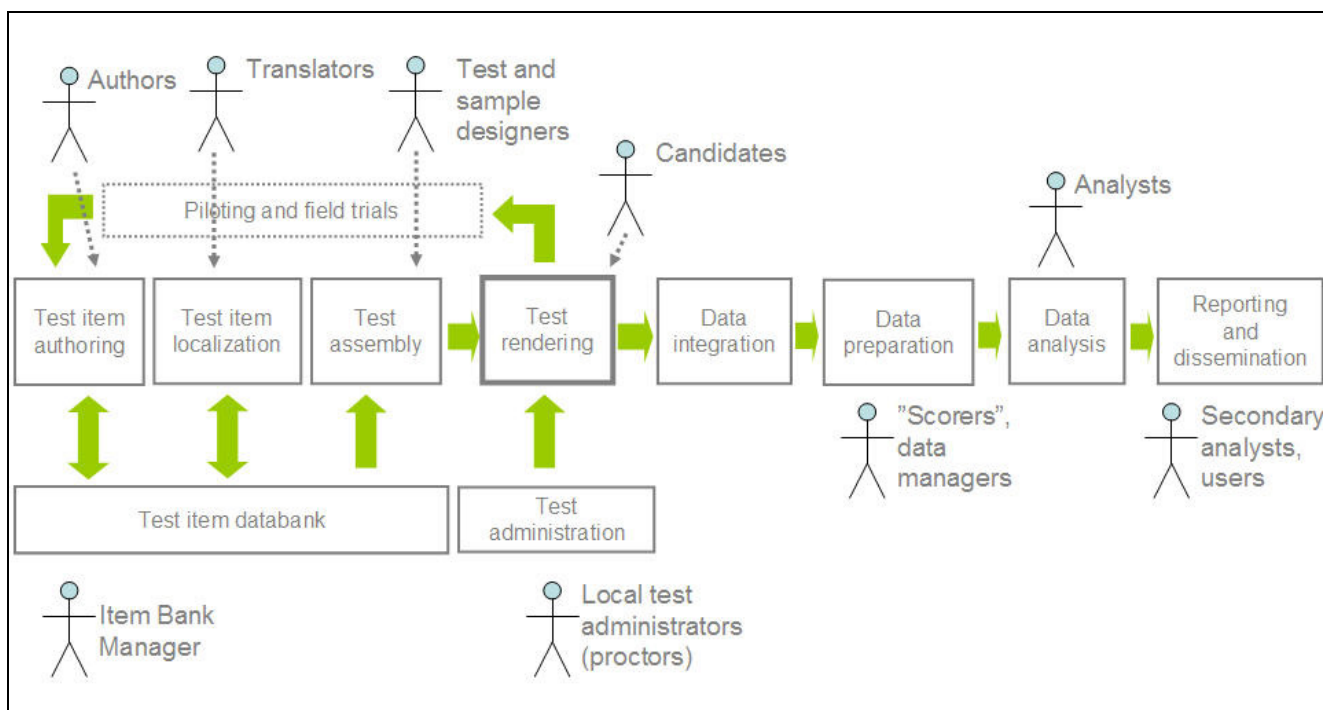


Figure 1: Stages and roles in the design and delivery of the survey

In terms of functionality, the following tools and components are needed:

- Test-item authoring, editing and preview functionality supporting a system of distributed authors scattered around Europe.
- Test-item databank functionality providing efficient storage, management and version control of test-items. This tool should also encourage visibility and sharing of resources between the various roles associated with the stages of the test-item life-cycle.
- Test-item translation functionality, supporting the localization of test-items, instructions and accompanying questionnaires to national languages.
- Test construction functionality, supporting the assembly of individual test-items into complete test sessions (compatible with the targeting and block substitution structure required by the overall design) as well as the allocation of students across tests at different levels.
- Test administration functionality supporting the management of respondents and test-sessions at the school level.
- Test rendering functionality supporting efficient and user-friendly presentation of tests-items to respondents as well as the capturing of their responses
- Data integration functionality supporting efficient assembly of response data coming from the participating schools.
- Data preparation functionality supporting all tasks related to the preparation of data files ready for analysis, including support for manual marking/scoring of open ended items.
- Data reporting functionality supporting online access to analytical results as well as download of files for statistical analysis.

## Architecture

The high level architecture of the software platform that has been designed to provide this functionality can be seen in Figure 2. The platform consists of a central Test-item databank interfacing three different tools over the Internet: 1) a Test-item authoring and editing tool, 2) a Translation management tool, and 3) a Test assembly tool. As a whole, these three distributed tools, plus the Test-item databank, are designed to support the central test development team in their efforts to develop and distribute the language tests.

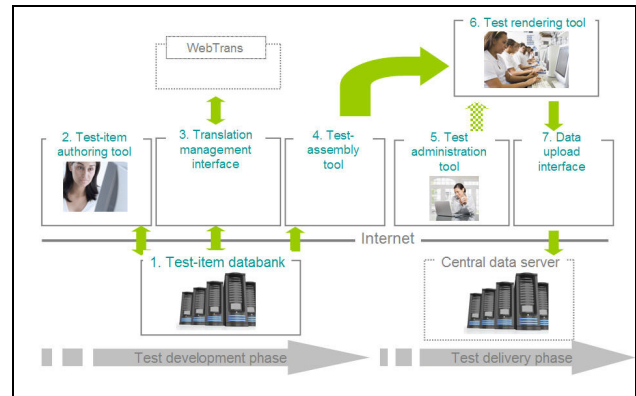


Figure 2: High level architecture

To support the test-delivery phase of the project, another set of tools will be provided. These are 1) a Test-rendering tool to be installed on the test computers in all the schools taking CB-testing and 2) a Test-administration tool supporting the various tasks of the local test administrators. Test rendering will take place on computers which are disconnected from the Internet. The Test-administration tool will however, provide an interface that will allow local test administrators to upload the collected data to a central database over the internet.

In the following paragraphs we will describe some of these tools in further detail.

## Test-item authoring

The test-items of the ESLC survey will be developed by an expert team of 50+ test-item authors distributed across Europe doing their work according to specifications and guidance provided by the central project team. Items will move through various stages of a predefined life-cycle including authoring, editing, vetting, adding of graphics and audio, pilot-testing, field-trial etc., each stage involving different tasks, roles and responsibilities.

The Test-item authoring tool is designed to support this distributed and fragmented development model. It is also designed to allow non-technical personnel to create tasks in an intuitive way by means of predefined templates for the various task-types that will be used in the survey. At any stage in the development, a task can be previewed and tested to allow the author to see how it will behave and look when rendered in a final test. The authoring tool will also support the capture and input of all the metadata elements associated with a task, including comments and descriptions, versioning metadata, test statistics etc.

The tool is implemented as a rich client by means of technologies like Adobe Flex and Adobe Air. This provides a very user-friendly and aesthetically pleasing environment for the various groups involved in the development of the tasks.

### Test-item databank

The Test-item databank is the hub of the central system providing long-term storage, version control and management of test-items and their associated metadata and rich media resources. Test-items will be uploaded to the databank by the test-authors to be seen and shared by others. When, as an example, a task has reached a stage in the development where an audio file should be added, the person responsible for this stage will download the task, read the audio transcript, create and attach the soundtrack and load the task back up to the databank. The databank will include a version control mechanisms keeping track of where the task is in the lifecycle as well as a secure role-based authentication system, making sure that only authorized personnel can see or change a task at the various stages in the life-cycle.

The Test-item databank is implemented in Java on top of Apache Tomcat and MySQL communicating with the various remote clients through Adobe Blaze DS.

### Translation management

It goes without saying that a software platform developed for foreign language testing will need to be genuinely multilingual. Not only will equivalent language tests be developed in the five target languages. User guides, menus, navigation elements and questionnaires will in addition be offered in all the national languages of the countries where the tests are taken. Each concrete test presented to a respondent will thus have two different languages; a target language and the national language of the location where the test takes place. This requires efficient language versioning and text string substitution support. It also requires an efficient, robust and scientifically sound translation management system.

Gallup Europe – one of the main partners of the Surveylang consortium - has already developed a translation management system called WebTrans for their large-scale international

survey operations, amongst other the Commissions' Flash Eurobarometer project. This WebTrans system supports central management of translators scattered all over Europe and a model of forward and back translation similar to the one that will be used for ESLC. The consortium has decided to make use of WebTrans and to create an interface between that system and the Test-Item Authoring tool.

### Test assembly

The Test assembly tool is without doubt the most sophisticated piece of software in the Surveylang platform. The tool is designed to support three important functions (see Fig. 3):

1. the assembly of individual test items into a high number of complete test sequences,
2. the allocation of students across these test sequences according to the principles and parameters of the predefined survey design (see above),
3. the production of the digital input to the computer-based Test rendering tool, and
4. the production of the digital documents that will be used to print the paper-based test booklets.

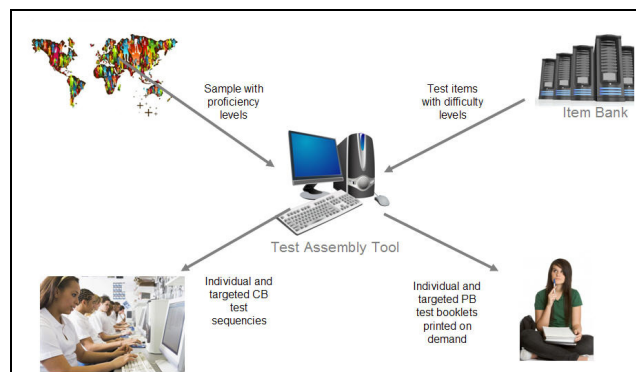


Figure 3: The roles of the test assembly tool

The assembly of test items into complete test sessions takes place in two steps. The first step is handled by the test designers and involves the construction of a high number of blocks (of approximately 15 minutes testing time). Each of these blocs contains tasks focusing on a single skill and is assigned one of four difficulty levels. The second step is fully automated and involves the assembly of blocks into test sequences, each containing two skill sections (of approximately 30 minutes testing time) plus a student questionnaire that will be administered to all the students. The difficulty level of the test sequences will directly derive from that of the

blocks which they are composed of. The test sequences are assembled according to the basic rules of the survey design which will be formalized in XML and interpreted by the Test Assembly tool. The tool will subsequently be able to handle changes to the design without any reprogramming.

In order to allocate students across test sequences, the Test assembly tool will also need to be able to make sense of detailed information about each single sampled student; especially the student's proficiency level (from the routing test), the target and national language of the student, the student's school and whether this school is taking a computer-based or a paper-based test. Based on this information and following the principles of targeted testing, the Assembly tool will randomly allocate students across the group of available test sequences matching their respective proficiency level.

The last task of the Test assembly tool is to produce the individualized test material for each single student, both for computer-based and paper-based testing. In the former case the material will be produced as a package of tests in a predefined XML-format ready for use by the Test rendering tool. In the latter case, the Test Assembly tool will produce a package of digital documents (PDF), each document containing the paper-based form of a named student. In both cases, digital packages will be produced for each single school in the sample based on the combination of target and national language relevant for that school.

What makes this an innovative approach is the fact that both modes (computer and paper) are served from the same source and the same system. This is reducing manual work and hopefully also manually-induced errors. According to the experiences of other international surveys involving complex test construction, printing of forms and the process of handing the right form to each individual student, are the most complicated and error-prone processes in the administration of the survey. By offering a solution where the package of individualized and named forms can be printed on demand for each single school, we hope to reduce these problems to a minimum.

## Test rendering

One of the critical challenges related to computer-based delivery of assessment tests is, in general, security. On the one hand, it is crucial that the content of the tests are protected from disclosure before and during the testing period. On the other hand, it is of utmost importance to create testing environments that are as equal as possible for everyone and where the students are protected from external influences of any sort (like access to the web, chatting channels, digital dictionaries etc.) while taking the tests. For the latter reason the test will have to take place on test computers that are disconnected from the net and where the desktop of available tools and software is fully controlled by the test administrators.

If the tests could have been taken on dedicated test computers brought into the schools for that very purpose, the problems would have been trivial. However, in a scenario where all tests will be taken on the schools' existing hardware platforms, this is more of a challenge. The solution that we are opting for is to boot the test computers from USB memory-sticks or CDs, including a minimum-size operating system (a Linux variety), the test rendering tool and the complete package of tests. In this way we will be in full control of the local desktop, we can block the access to Internet and we can monitor that the tests actually are taken in this closed and controlled environment. We are aware of many technical obstacles that will have to be overcome for this to happen smoothly in every technical environment. We will however use pilot testing and the upcoming field-trial to map and develop workarounds for these obstacles.

The test rendering tool will be implemented in Adobe Flex and run in the Adobe AIR runtime environment. It is designed to support the rich multimedia test format generated from the Test assembly tool.

The administration of the tests at school level will be managed through the Test administration tool. This tool is also including the functionality to upload the captured data to the central data server. In order to reduce the requirements for the local hardware platform to a minimum, this part of the Surveylang software platform will not be dependent on a local area network (LAN). It is expected that this approach will increase the number of schools that are in a position to take the computer-based version of the tests. As a

side effect, we also expect a lower number of failures due to local technical problems.

## Open Source

The SurveyLang software platform will be developed as open source. This means that the platform will be free and open for use and extension by others. As soon as the development has reached a certain level of stability, the code and documentation will be made public and distributed under a standard open source license. The project is also based on open source development strategies, including a will to mobilize external developers and testers and to collaborate with other projects. As much as possible the development is based on existing open source frameworks, tools and components. We are also following open standards and using technologies that support an open source development model.

## Conclusions

The software platform currently being built by the SurveyLang consortium is designed from the bottom up to meet the complex requirements deriving from large-scale international assessment surveys. This includes support for the standard questionnaires that are normally part of these surveys. The software is built to support language testing, but could easily be extended to support other task types and subject domains. It is built around a generic task

structure and life-cycle model that maps to international standards like QTI and DDI.

The implementation of the software started in September 2008 so there is still a way to go before the first versions of the software will be released. A complete suite including all the tools described in this article will be ready for the ESLC field trial in the first quarter of 2010.

## The author:

Jostein Ryssevik

E-Mail: [Jostein.Ryssevik@ideas2evidence.com](mailto:Jostein.Ryssevik@ideas2evidence.com)

Phone: (+47) 91817197

Web: [www.ideas2evidence.com](http://www.ideas2evidence.com)

Jostein Ryssevik has twenty-five years of experience in social science data management. He has headed a long range of international software development projects within the fields of statistics, data analysis and knowledge management. From 2001 to 2005 he served as technical director of the UK-based software company Nesstar Limited developing tools for Web-based publishing and analysis of statistical data. Ryssevik is an expert on statistical metadata. He has played an active role in the development of the DDI metadata standard and is one of the founders and managers of the Open Data Foundation, a non-profit organization dedicated to the adoption of global metadata standards and the development of open-source solutions promoting the use of statistical data. He is currently managing the independent consultancy company ideas2evidence and is responsible for the development of the software platform for the European Survey on Language Competences.