



United Nations  
Educational, Scientific and  
Cultural Organization



## REQUEST FOR INFORMATION (RFI)

### UNESCO Institute for Statistics

June 2019

## 1. PURPOSE OF REQUEST FOR INFORMATION

As custodian agency for SDG4, the UIS is launching this Request for Information to identify qualified institutions (consultancy firms, universities, etc.) for the long-term provision of item banking and exchange platform with focus on serving the tracking and monitoring to global and thematic indicators related to learning in SDG4. The sought-after institution(s) should have strong technical expertise in one or more of the following areas:

- 1) Item Banking
- 2) Data processing, analysis and reporting for large-scale learning assessment
- 3) Experience developing web applications that perform real-time psychometric data analysis, including production of diagnostic score reports and estimation of Item Response Theory item parameters.
- 4) Relevant experience in the development or customization of assessment software to suit the operational needs of clients responsible for administering assessments and reporting results.

The development of generic software not customized for specific client needs will not be considered.

## 2. BACKGROUND

The Sustainable Development Goals (SDGs) set out new ambitions for education, with SDG 4 requiring a quality education from pre-primary to upper secondary level of education for every child by 2030. The focus on quality education is a deeper and more demanding focus than the Millennium Development Goals (MDGs), since it puts more emphasis on learning outcomes.

SDG target 4.1 calls on all UN Member States to 'ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes' by 2030. Specifically, indicator 4.1.1 measures 'Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex'.

Around the world, there are dozens of countries for which we do not know the levels of learning and the costs of assessing them are very high. There is also a need for better technical documentation to guide countries in producing proficiency statistics. The challenge is to find the most cost-efficient, fit-for-purpose way of doing this. A practical low-cost solution is an item-bank and test-scoring Platform that utilizes machine learning algorithms to lower costs and increase efficiency of large-scale assessments. Specifically, the use of algorithms developed for high-dimensional analysis of natural language, familiarly used to generate web search results, machine translations, and shopping recommendations, allows the Platform to automate many frequent tasks that would otherwise



require manual input. Manual activity will be required where it is most useful, for one-time activities related to initial item classification and definition of performance level descriptors, while routine tasks, such as item response scoring, test form development, item response analysis, test score interpretation, standard setting, and test scale linking, may be largely automated or heavily machine-assisted. Recent evidence from machine learning research indicates that trained, cross-validated neural networks are comparable or superior to human judgment and traditional statistical models for both prediction and classification. The item bank focuses on the performance level descriptors and the exact cognitive requirements of each item, which, as the basis for scoring and data analysis, would facilitate expert-level automation.

The recent finalization of the Global Content Framework (GCF) for Reading and Mathematics provides a solution for indexing current and future assessments. The common framework simplifies the use of dimensional mapping algorithms to link assessment content, performance level descriptors, and student assessment results.

### **3. PROJECT DESCRIPTION**

#### **General objective**

The purpose of the Item-bank Project is to create an objective and highly reliable reading and mathematics assessment scoring and reporting tool, is free and available to children all over the world.

The item bank and test platform would allow assembly and scoring of fit-for-purpose assessments for countries' own tracking and for reporting on SDG 4.1.1 grades 2/3, the end of primary and end of lower secondary. The statistical focus of the UIS item bank is on the performance level descriptors (PLDs) and the exact cognitive requirements of each item as the basis for scoring and data analysis (in contrast to the most common current practice, which uses the cognitive definitions to specify test design, but largely ignores them in subsequent data analysis and scoring).

#### **Specific Objectives**

The specific objectives are to build a Platform that facilitates three functions

- ✓ Item Banking
- ✓ Test assembly, scoring and analysis with the rules that allow reporting on SDG4.1.1 in reading and math.
- ✓ Linking country-specific test results to international PLDs and the GCF.

### **4. SCOPE OF WORK: THE SDG4 TEST PLATFORM**

The SDG4 item and test Platform is the main deliverable based on the work carried out by the UIS regarding the item bank Platform design and in close collaboration with countries and other partners as part of the broader objective of creating cost effective ways of testing and reporting for SDG4. The



final output is a single-server application specifically to facilitate the assessment needs of countries through their in-charge public agencies (such as Ministries of Education).

Concretely, these are the features the UIS requires for the Item Bank and Test Platform (the Platform)

### 3.1. Management

One or more UIS administrators must be able to control high-level user access to the UIS Item Bank. Different countries require their administrative and operational management structures to be mirrored in the UIS Item Bank so that country managers can control access to the UIS Item Bank of lower level users who perform operational tasks, such as item development, item classification, test development, and analysis. Each user should be assigned one or more roles, corresponding to the system functionality that is required for them to meet their responsibilities.

The organization structures of one country should not interact or interfere with the structures of another country. However, users with management or operational roles in one organization (country) may have management or operational roles in another organization. Users' roles in one organization are unrelated to any privilege or access to data or functionality in the context of any other organization.

- ✓ Globalization / Localization
- ✓ Scalability
- ✓ API

### 3.2. Content: Item Banking

- i. Items may be created online or uploaded to the Platform. Uploaded item definitions must be loaded in a standard schema and data format recognized by the Platform (e.g., QTI<sup>1</sup> objects in JSON or XML format). NOTE: strict QTI interoperability is not essential, as many target countries may not be using QTI-compliant item banks; minimum required item interoperability must be able to import multiple choice items with a variable number of response options using a combination of formatted or unformatted text and one or more images from character-delimited or JSON formats.
- ii. Items formats must support different delivery modes: paper and pencil, online, offline, and interview. Online item format support multi-state items that modify the information presented to and collected from test takers in response to timed events or test taker interactions. All online item formats are available in offline delivery using the offline test delivery mode of the Platform.
- iii. Items must allow a variety of data capture and scoring methods. Minimum required response formats include: single and multiple choice, text and numeric short answer, speech, drawing, extended text response, and document highlighting.

---

<sup>1</sup> IMS Question & Test Interoperability, <http://www.imsglobal.org/question/index.html>



- iv. Item definitions for all types of items must allow online editing. Item definitions must allow specification of scoring guides for automated machine-scoring and manual scoring for both finite-response and open-ended response. Item definitions must separately define each distinct response category.
- v. Organizations/countries may use the item classification interfaces to classify items against organization-specific curricula or assessment frameworks.

### 3.3. Test Development

- i. The Platform provides tools for statistical and semantic analysis of proficiency levels descriptors.
- ii. Organizations/countries may use the PLD definition interfaces to define organization-specific proficiency levels that map to organization-specific curriculum.
- iii. On request and where sufficient data exist, the Platform will perform semantic and statistical analyses that map organization-specific proficiency levels to the GCF-mapped PLDs to determine how they compare to international interpretations.
  - i. Test design tools allow users to target one or more PLD from both reading and mathematics and specify other constraints, which may include number of items, testing language, and delivery mode.
  - ii. The Platform provides test design templates that correspond to common testing situations (e.g., single-form, multi-stage adaptive, instruction-drill-and-test learning modules). The Platform distinguishes between logical elements of the test design (i.e., testlets or modules) and the items that populate these elements. Users can modify and extend the templates to suit their assessment needs. Test developers populate the test design elements with items drawn from the item bank using manual or automated item selection and ordering.
  - iii. Automated test assembly functions select items and create a test form (or test schema, for tests delivered on the Platform) that can be downloaded or accessed online. Automated item selection will optimize selection against existing item family and item enemy definitions. Manual item selection uses search fields and test design constraints to find and filter items by item metadata and item content.
- iv. Test definitions may be copied and modified to create new tests.
- v. Users explicitly define testing constraints for Platform-delivered testing such as testing time, length of testing window, multiple sessions, secured testing, navigation between items, visibility of test progress, language switching, and review page accessibility.

### 3.4. Test Taker Management

- i. The Platform's primary purpose is to monitor country-level progress with respect to SDG4. The Platform supports offline delivery of tests using a secured test delivery application installed on a



mobile device. The mobile application must be integrated with the Platform for synchronization of data, functionality, and security. The mobile application is compatible with current Windows mobile devices.

- ii. Online and offline test administration save response data in real time, allowing test takers to recover interrupted testing sessions with their existing response data and progress.
- iii. Where required by the test definition, Platform-based testing (online and offline) is secured by testing-site-specific or test-administrator-specific authentication, which is authorized on a per-test basis by the organization.
- iv. Any test may be administered using non-system modes (e.g., paper-and-pencil), and the response data may be uploaded to the Platform for scoring and analysis. Uploaded test response data must include a PSU identifier that has been provided by the Platform in a sample definition that pre-exists in the Platform.

### **3.5. Test Response Data Analysis**

- i. The Platform provides functionality to manage and collect data from standard setting panels using item-based methods (latent class, Angoff-style and Bookmark) and population-based methods (contrasting groups). Standard setting output includes numeric thresholds and sets of classified items for use in defining described scales. The numeric thresholds estimated from standard setting procedures are available for use in the definition of reports.
- ii. The Platform provides methods for linking items and scales of different test analyses. Required methods include common-item linking using the Stocking-Lord and Mean-Mean IRT methods, test-score linking using equipercentile and nonlinear regression methods.

### **3.6. Test reporting**

- i. The Platform estimates test taker scores for each testing domain using user-supplied specifications. The Platform calculates the following types of score estimates: linear equation scores (sum, average, factor), weighted maximum likelihood (IRT), expected a posteriori (IRT), plausible values (IRT), manifold (GCF mapping). For IRT-based methods, users may specify the use of conditioning variables, which may include other testing domains, for the estimation of each scoring domain.
- ii. The Platform supports definition of data-only reports, which may include data for estimated scores and individual items. Data-only reports are produced in delimited data tables, where each column corresponds to a single variable, and each row corresponds to a single observation, test taker, or aggregation unit. Data-only reports may include sample-design data to facilitate in-depth secondary analysis in third party software.
- iii. The Platform supports formatted reports which can include a combination of text, graphics and tables that incorporate statistical estimates and their standard errors. Report content may be



static, appearing the same for all versions of a report, or dynamic, changing according to the specific values of statistical data used to generate the report.

## 5. PERIOD OF ASSIGNMENT

The implementation of the Platform coincides with ongoing activities related to SDG4. These activities include the collection of test items from primary sources, indexing of test items against the GCF, and data collection activities of stakeholders. The complete functionality must be implemented within 12 months since the signature of contract.

It is expected the firm/entity to be able to provide a perpetual, worldwide, royalty-free license of the assessment software application to the UIS.

## 6. SCOPE OF THE REQUEST FOR INFORMATION

The current Request for Information (RFI) requests the suppliers or other players in the market to provide information about available products, technologies, services, qualifications and experience in order to gain required knowledge before finalizing the solicitation documents and technical specifications/TOR. Request for Information (RFI) is therefore a mechanism of gathering information and will not constitute a formal procurement process.

Vendors/legal entities are strongly encouraged to respond to this inquiry, as this information will be instrumental in preparation of the procurement strategy and finalization of the solicitation documents.

Your submission must be received by UNESCO no later than **3 July 2019** on email address: **uis.procurement@unesco.org** (max file size 10MB). Please enter **SDG4 Test Platform** in the subject line.

This request is strictly a mechanism for gathering information and will not constitute a procurement process. Submitting a reply to a Request for Information does not automatically guarantee receipt of the solicitation documents when issued. Invitations to Bid or Requests for Proposals and any subsequent purchase order or contract will be issued in accordance with the rules and procedures of UNESCO.

For further information please contact: **uis.procurement@unesco.org**


**ANNEX 1:** [to form part of your submission to this Request for Information – RFI]

<b>1. VENDOR INFORMATION</b>	
UNGM ID Number (if available):	
Legal Company Name:	
Company Contact Person:	
Address:	
City:	
State:	
Country :	
Zip / Postal Code:	
Telephone :	
Email:	
Company Website:	
<b>2. Presentation of your enterprise including number of staff, turnover, years in business:</b>	
<i>[please insert or attach as an annex]</i>	
<b>3. Previous experience concerning item banking; data processing, analysis and reporting for large-scale learning assessment; experience developing web applications that perform real-time psychometric data analysis, including production of diagnostic score reports and estimation of Item Response Theory item parameters; relevant experience in the development or customization of assessment software to suit the operational needs of clients responsible for administering assessments and reporting results.</b>	
<i>[please insert or attach as an annex]</i>	
<b>4. Reference list demonstrating your qualifications for participating in possible upcoming solicitation process; at least (02) references:</b>	
<i>[please insert or attach as an annex]</i>	
<b>5. ADDITIONAL INFORMATION:</b>	
<i>[please insert or attach as an annex]</i>	