**Annex 1: Pilot instrument of the UNESCO General Education Quality Analysis/Diagnosis Framework (GEQAF)**

**Analytic Tool, Assessment**

**Paramount Question:**  **how assessments can contribute to improving the quality of our education system and learning effectiveness?**

|  |  |  |
| --- | --- | --- |
| **Diagnostic question** | **A brief summary of responses to the diagnostic question from a Team of national education policy makers, planners, managers and experts conducting the diagnosis and analysis**  | **Priority actions and knowledge gaps identified in the process of addressing the diagnostic question** |
| **Assessment policies, frameworks and methods** |
| 1. Do we have a national strategy / policy / position paper on educational assessment? If yes, how recent is this? Which educational levels (both in terms of International Standard Classification of Education (ISCED) and in terms of location (local – regional – national) and subjects are covered by this? Has it been evaluated?
 |  |  |
| 1. To what extent is the choice of purposes, targets and subject matters for assessment, for example in national assessments, related directly to what the country thinks of as important in terms of learning outcomes for its learners and *not* only in terms of what is easy to assess? (Technical note VII.2) [**Analytical Tool on Competencies**] [**Analytical Tool on Curriculum**]
 |  |  |
| 1. What have been the criteria used to determine the coverage of the assessment and the level at which national assessments are conducted? Are these criteria linked to clear objectives and goals of the assessment? Is there evidence that the coverage and the levels at which the assessments are made contributed to improvement of education system quality?
 |  |  |
| 1. In general, to what extent is assessment in this country effective? To what ends? Is it inclusive? In what way? What evidence do we have for this? Do we know where the system stands in terms of achievement outcomes at every level? **[Analytical Tool on Equity and Inclusion]**
 |  |  |
| **Implementation of assessment** |
| 1. If there is an educational assessment policy has it been implemented / enacted? How do we know? At what levels is assessment implemented? What are the objectives of this?
 |  |  |
| 1. Is there evidence that the implementation of the assessments is according to rules of good practice, incl. inclusiveness? What is this based on? [**Analytical Tool on Equity and Inclusion**]
 |  |  |
| 1. Who implements assessments? How does this vary by types of assessment?
 |  |  |
| 1. How are tests conceptualised (i.e. how are test items developed) and what is the conceptual basis for this (for example, a curriculum / syllabus analysis or rather an orientation of ‘life skills’)? What psychometric methods and techniques are used to classify items[[1]](#footnote-1), and to what extent are these item characteristics taken into account in the development of achievement tests? Are open and closed items used? In terms of test conceptualisation, is there a good mix of standardised and non-standardised testing available? (Technical note VII.3, Alternative assessment)
 |  |  |
| 1. Are assessments also measuring ‘associated factors’ that facilitate analysis (e.g. looking at age, gender, socioeconomic status and other background information)? [**Analytical Tool on Equity and Inclusion**
 |  |  |
| 1. If applicable, how are data processed and fed into a centralised information system?
 |  |  |
| 1. What is the evidence that participation in international quality assessment (LLECE, PISA, SACMEQ and others) help us to bench mark the quality of our education system? What has been our and others experience of international assessments? If we have not participated, was it a deliberate decision and, if so, why?
 |  |  |
| **Utilisation of assessment results** |
| 1. What mechanisms do we have for making the evaluation of the assessment results inform education policy and practice (at classroom, school, regional and national level)? How often do we use these mechanisms? What is the evidence that we do such evaluation in a purposeful and systematic way? [**Analytical Tool on Relevance**] [**Analytical Tool on Governance**]
 |  |  |
| 1. How do we interpret the findings from evaluations of assessment results findings, and how do we make sure that educational assessments have the intended impact of improving the education system quality and learning effectiveness? How do we communicate our evaluation so as to focus on how we can do better? How are outcomes data linked to other variables, such as finance data, which permits rigorous analyses? [**Analytical Tool on Financing**]
 |  |  |
| 1. Are assessment results made public, and to whom (for example, individual student results to parents / carers; school rankings to the general public, etc.)? [**Analytical Tool on Governance**]
 |  |  |

The diagnosis and analysis above should culminate into identifying critical problems requiring urgent attention and the necessary information and knowledge for addressing them. It is also necessary to clearly formulate action plan and clear identification of roles and responsibilities and timelines as well as required human, financial and organizational resources which the action plan might entail. At this stage it is a question of prioritizing the priorities and knowledge gaps identified in the right most column of the table above to focus action on those areas severely hampering progress.

|  |
| --- |
| **Priorities for action (Assessment)** |
| 1. What are the key areas to be addressed urgently to make assessment contribute to the quality of our education system?
 |  |
| 1. What are the knowledge gaps which need to be filled for an evidence-based policy and practice of school-based and national assessments?
 |  |
| 1. What are the required actions to deal with the priority constraints and the identified knowledge gaps?
 |  |

**Annex 2: Format for feedback on the piloting of the individual Analytic Tool of GEQAF**

***To be completed at the end of the discussion of each Analytic Tool***

|  |
| --- |
| **Analytic Tool: Assessment** |
| 1. Which questions did you find unclear or hard to understand? If so how would you suggest they be reformulated?
 |  |
| 1. Which of the questions did you find less relevant in your context? Why?
 |  |
| 1. Which questions of critical importance in your context are missing in the toolkits?
 |  |
| 1. Which questions did you find too demanding on data and information relative to the significance of the issue for ensuring quality education?
 |  |
| 1. Would you have preferred more and detailed question or were the set of questions in the toolkit adequate to discuss the issues in depth?
 |  |
| 1. To what extent did this toolkit help you analyze the issues raised comprehensively?
 |  |
| 1. What kind of further support materials you would have needed for a more in-depth analysis?
 |  |
| 1. How much time was allocated for the discussion of this toolkit? Would it have required more or less time and if so how much?
 |  |
| 1. Would you use this toolkit in the future? Is so, how often?
 |  |

**Annex 3: Summative evaluation of GEQAF and the guidelines for piloting**

To be completed by the pilot Core Team with inputs from Heads of Departments and/or agencies

|  |
| --- |
| **The procedure of implementation** |
| 1. What significant adjustments did you make to the procedure suggested for piloting by UNESCO and why?
 |  |
| 1. What further improvements to the UNESCO guideline and piloting instrument would you suggest?
 |  |
| 1. To what extent do you think the results from applying the UNESCO education quality framework have been worth the time and resources you have invested in the exercise?
 |  |
| 1. Do you think you would use the framework (or parts of it) from time to time to check the pulse of your education system? If so, how often?
 |  |
| 1. What next steps were agreed or proposed to address major challenges identified during the diagnostic exercise?
 |  |
| 1. Who will be responsible and for what in following up on actions agreed or proposed
 |  |

1. Main psychometric item characteristics include difficulty and discrimination, but there exist other characteristics. Item difficulty is often established using ‘scaling’ according to the Research methodology. The process for doing this is called ‘item calibration’. New models for scaling, such as 2PL and 3PL, are gaining in significance. [↑](#footnote-ref-1)