



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



GLOBAL
ALLIANCE
TO MONITOR
LEARNING



SUSTAINABLE
DEVELOPMENT
GOALS

Procedural Alignment Tool

GAML Fifth Meeting
17-18 October 2018
Hamburg, Germany



Contents

INTRODUCTION: PROCEDURAL ALIGNMENT TOOL	3
PROCEDURAL ALIGNMENT QUESTIONNAIRE	5
SCORING RULES – PROCEDURAL ALIGNMENT TOOL	12

Introduction: Procedural Alignment Tool

The UNESCO Institute for Statistics (UIS) has developed several self-report questionnaires that countries will complete when they submit their locally developed national learning assessment results to UIS to use as part of a country's documentation demonstrating progress in attaining Sustainable Development Goal 4 (SDG4). Currently UIS has one questionnaire related to procedural alignment that a country's representatives will complete using UIS's DART data collection system: Alignment of the National Assessment's Procedures for Assessment Development with the UIS's guidelines, *Principles of Good Practice in Learning Assessment*. A country's responses to this questionnaire will help UIS to determine its eligibility to use its locally developed national assessments to support its progress for achieving SDG4. The questionnaires are designed in such a way that respondents can complete them in a short time without a lot of writing.

To support the use of the results of its national assessment as measures of progress in achieving SDG4, a country's test development procedures should follow internationally recognized good practices of test development and data collection. These practices go beyond content coverage: The good practices related to how the test was developed, and how the data were collected and used in the assessment development. To help countries organize and monitor their development of their local national assessment of learning, UIS created guidelines, which are contained in its publication, *Principles of Good Practice in Learning Assessment*.

Based on the principles describes in this document, UIS developed a questionnaire that a country's representative will use to report the procedures used to develop of the assessment in each of these 12 important areas:

- The quality of the assessment team responsible for implementing the learning assessment.
- The use of technical standards to guide assessment activities.
- The use of an assessment framework to guide test development.
- The procedures used to develop test items.
- The linguistic quality control during writing the test.
- Following appropriate procedures for designing cognitive instruments.
- The quality of the sampling of students.
- The standardization of the operational administration of assessments.
- The quality of managing data.
- The quality of equating scores from one year to the next.
- The quality of the analyses of assessment data.
- How well the results are reported and used.

The procedural alignment questionnaire has specific questions within each of these 12 areas so that a country may describe its alignment to good practices in these areas.

The results from these UIS questionnaires not only provide information to UIS about the appropriateness of a country's national assessment for supporting achievement of SDG4, but they provide feedback to each country about its national assessment. The feedback from UIS will include information about which areas of test development a country should focus to improve the quality of its local national assessment.

Procedural Alignment Questionnaire

INSTRUCTIONS

Dear Colleagues,

The UNESCO Institute for Statistics (UIS) thanks you for helping us to obtain information about country's your national learning assessment. This questionnaire ask about the components and procedures sometimes used to design and develop national assessments of learning outcomes.

Your participation in completing this questionnaire is very important. Your responses will help UIS validate that your country's national assessment of student learning outcomes can be used to measure your country's progress in achieving Sustainable Development Goal 4.

BEFORE STARTING TO ANSWER THE QUESTIONS

- ✓ Become familiar with the layout of the questionnaire. The questionnaire asks you questions related to the procedures your country used when developing your national assessment. The questions are organized around these components: assessment team, technical standards guiding activities, assessment framework, development of items, linguistic quality control, designing instruments, sampling, standardized operations, managing data, equating scores, analyses of data, and reporting results.
- ✓ Identify the assessment person(s) who could help you answer the questions. Prepare a list and contact information for the key persons you can call to obtain or to verify information so you can answer each question.
- ✓ Identify the educational level for which your county is submitting its national assessment of learning outcomes to support Sustainable Development Goals 4. For which level is the national assessment measuring: primary education, lower secondary education, or upper secondary education?
- ✓ Identify the official name and educational level of the national assessment about which you are answering the questions. You should be answering questions about only this national assessment, even though your country my have several assessment programs.
- ✓ The format of every question asks you to respond "yes" or "no". Read each question and decide whether your country's national assessment contains the component or uses the procedure. Answer the question "YES" if your national assessment program contains the component or uses the procedure. Answer the question "NO" if your program does not contain the component or does not use the procedure.
- ✓ Some of the questions will ask if there is a document that contains certain information. Be sure to obtain a copy of the document ahead of time so you can answer the question posed.
- ✓ There is a glossary of the common terms used in assessment that are contained in the questions. Please refer to this glossary frequently to help answer the questions.
- ✓ Be sure to answer every question. If you do not know the answer, please make a sincere effort to find the answer. If you cannot find the answer, please respond "NO" to the question.

Glossary of Technical Terms appearing in the Questionnaire

Census	All students in the target population are tested.
Clean the data	Reviewing all numerical data carefully and systematically to verify correctness, detect and remove or correct inaccuracies, and correct or remove incomplete or irrelevant values.
Framework /blueprint	A table describing the major content categories, skills, and content standards that a test should assess. The table describes the percentage of questions the test should contain for each content-skills combination included.
Operational version of the test	The final version of a test that is administered to students to assess learning. This is contrasted with field test versions of the test, or preliminary versions of a test for which students' scores do not matter to the final assessment process.
Pilot/field the tests	Administering the test questions to a sample of students and collecting data and other information about whether the questions are functioning as intended and using that information to select or improve the questions. Sometimes called tryout tests.
Probability sampling methodology	Any of several methods of selecting persons to be in a sample that is random and gives any member of the population an equal chance of being selected. A non-probability sample does not use a random selection process but uses other methods such as judgment or convenience. Among the probability sampling methods are: simple random sampling, systematic sampling, stratified random sampling, cluster random sampling, multi-stage random sampling, etc.
Quality that the data from the test must achieve	High quality data are accurate, do not contain false values, are complete with no missing values, conform to established formats, do not contain repeated or duplicated values, etc. A document can be created that describes the quality that the data from the national assessment program must meet.
Reliability coefficients	Any of several statistical indices that quantifies the amount of consistency in assessment scores. Reliability is the amount of consistency of assessment results.
Sample of students	A subset of students who have been selected from the population of all students.
Sampling frame	A list of all persons who are eligible to be included in a sample. For example, a list of all students in grade 6 of primary school in the country. Samples of students may be drawn from this sample frame.
Secure facility	A physical place in which test materials are kept and in which testing staff work that prevents non-authorized persons from seeing or obtaining copies of the testing materials.
Stakeholders	Persons or groups with an interest in the results of an assessment, usually because they will be affected by decisions made about them using the test results.
Standard error of a statistical estimate	When sample data are used to calculate a statistic to estimate a population value (e.g., the sample average score is used to estimate the population average score), the sample estimate is usually not exactly equal to the population value. The standard error of a statistic is an index that estimates the standard deviation of many sample estimates of the same population value. It is, therefore, used to describe the expected accuracy of the sample estimate.
Standard error of measurement	A statistical index that estimates the standard deviation or spread of a hypothetical distribution of error scores. Scores students receive on a test are called obtained scores. These obtained scores contain errors of measurement. The standard error of measurement is used to estimate the average magnitude of these errors.

Standardization	When the procedures, administration materials, scoring rules, are fixed so that, as far as possible, the assessment is the same at different times and places.
Statistical design	The plan one develops to assure that the data collected and the statistical analyses done with the data will be accurate.
Students with disabilities	Students that have one or more of the following: autism, deafness, blindness, deaf-blindness, developmental delay, emotional disturbance, hearing impairment, orthopedic or muscular impairment, health impairment, specific learning disabilities, or intellectual impairment.
Sub-populations	All those students who are in a defined category of the target population. Examples of sub-populations include male students, female students, students with disabilities, urban students, and rural students, etc.
Summary report of the results	After a national assessment of student learning is administered the results are analyzed and summary statistics are compiled. The summaries are organized, interpreted, and included in a comprehensive report.
Target population	All persons that you would want to assess if resources permitted it. For example, all students enrolled in grade 6 of primary school in the country. A sampling frame may include everyone in the target population, but it might not. For example, a list of all students in grade 6 may not include students who dropped out, it might not include students attending grade six who officials have forgotten to register with authorities, etc.
Technical procedures for equivalence of test	Tests vary from year to year since questions differ, or the questions on a test in one year may be easier or more difficult than the questions on the test in another year. Technical (statistical) procedures exist to adjust the scores from each year's test so they are equivalent. Equivalent means that the test measures the same skills and abilities with the same degree of accuracy. Statistical procedures such as equating methodology is employed to assure equivalence.
Valid	Validity is the soundness of one's interpretation of the assessment results.

Country information

Country for which this questionnaire is being completed: _____

Date of data collection: _____

Respondent contact information

Country: _____

Name: _____

Affiliated institution: _____

Job title: _____

E-mail: _____

Telephone number: _____

Please provide information on all key individuals who contributed to answering this questionnaire. For each individual, please provide (1) name, (2) affiliated institution, (3) job title, and (4) contact information (e-mail and telephone).

Procedural Alignment Questionnaire Version 2.2

Official name of the national learning assessment that you are answering these questions:			
Educational level assessed by the national assessment for which this questionnaire is being answered:	PRIMARY	LOWER SECONDARY	UPPER SECONDARY

Assessment Team Responsible for Implementing the Learning Assessment

1	Have at least 70% of the members of the official team/group that developed the national learning assessment been trained in how to develop a national learning assessment?	YES	NO
2	Do you have a <i>secure facility</i> for housing the team/group and testing materials?	YES	NO

Technical Standards to Guide Assessment Activities

3	Do you have an official document that describes the <i>statistical design</i> of how the students will be sampled and administered the assessment?	YES	NO
4	Do you have an official document that describes the <i>quality that the data from the test must achieve</i> in order to be <i>valid</i> ?	YES	NO

Assessment Framework

5	Did you use a committee of subject matter and educational experts to create a <i>framework /blueprint</i> that guided the test question developers?	YES	NO
---	---	-----	----

Development of Items

6	Were the draft questions been reviewed and evaluated by relevant persons outside of the team/group that created them before those questions were <i>piloted/field-tested</i> ?	YES	NO
7	Were all of the questions appearing on the <i>operational version</i> of the national assessment <i>piloted/field-tested</i> ?	YES	NO

Linguistic Quality Control

	Did you implement a translated version of your national learning assessment?	YES	NO	UIS do not score
	IF YOUR ANSWER IS <u>NO</u> , PLEASE JUMP AHEAD TO QUESTION 10			
8	Were the persons translating the national learning assessment instrument(s) trained on how to translate and adapt these tools for the target population?	YES	NO	
9	Were the translated assessments and instrument(s) evaluated and improved on the basis of <i>pilot/field tests</i> ?	YES	NO	

10	Did you implement any testing accommodations to assure that <i>students with disabilities</i> could appropriately demonstrate their knowledge and skills?	YES	NO
----	---	-----	----

Designing Cognitive Instruments

11	Did you conducted a formal study of how well the national learning assessment questions are aligned with the officially establish test framework/blueprint?	YES	NO
----	---	-----	----

Sampling

12	Do you have a formal document that clearly defines the <i>target population</i> for which the results of national learning assessment are reported?	YES	NO
13	Do you have a formal document that clearly defines each of the <i>sub-populations</i> for which the results of the national learning assessment are reported?	YES	NO
	Is your national assessment a <u>census</u> (i.e., all students in the target population are tested)?	YES	NO
IF YOUR ANSWER TO QUESTION IS <u>YES</u> , PLEASE JUMP AHEAD TO QUESTION 16			
If your national learning assessment based on a <i>sample of students</i> ,			
14	Do you have a document describing the <i>sampling frame</i> for the target population(s)?	YES	NO
15	Have you used a <i>probability sampling methodology</i> ?	YES	NO
16	Do you report <i>standard errors</i> of you results?	YES	NO
17	Did you conducted formal analyzes to verify that the students you tested is representative of the target population(s)?	YES	NO

UIS do not score

Standardized Operational Administration of Assessments

18	Did you conduct training on <i>standardization</i> for those persons who handled the assessment materials and implemented the national assessment?	YES	NO
19	Did you monitor and control the distribution of testing materials so that the integrity of the testing results would be maintained?	YES	NO
20	Did you account for all testing materials before and after the administration of the national assessment?	YES	NO

Managing Data

21	Do you have a written plan for scoring the national assessment and for maintaining the accuracy of the scores?	YES	NO
22	Did you follow appropriate procedures to <i>clean the data</i> (verify correctness and detect and remove/correct inaccuracies, correct/remove incomplete or irrelevant) before processing the national assessment results?	YES	NO

Equating Scores

23	Do you have a formal document describing the <i>technical procedures</i> to assure that each year’s national learning assessment scores are equivalent to the previous year’s assessment scores?	YES	NO
24	Did you implement the technical procedures to equate the national learning assessment scores with previous years?	YES	NO

Analyses of Assessment Data

25	Did the analysis of your national assessment results follow a formally defined <i>data analysis plan</i> ?	YES	NO
26	Do your data analyses estimate and report average scores for the entire target populations?	YES	NO
27	Do your data analyses estimate and report average scores for <i>sub-populations</i> ?	YES	NO
28	Do you report <i>reliability coefficients</i> or <i>standard errors of measurement</i> for your assessment?	YES	NO

Reporting and Using Results

29	Do you prepare a <i>summary report of the results</i> of the national assessment each time it is administered?	YES	NO
30	Do you have a document describing a plan for monitoring how national assessment data are being used over time?	YES	NO

Comments:

Scoring Rules – Procedural Alignment Tool

Procedural Questionnaire Category	Num. Quest.	Grouping Name	Max. Score	Sufficient Score
1. Assessment Team Capacity	2	Capacity and Technical Standards	4	3 or more
2. Technical Standards	2			
3. Assessment Framework	1	Instrument Development	4	3 or more
4. Development of Items	2			
5. Designing Cognitive Instruments	1			
6. Linguist Controls	1 or 3	Linguistic Control ^{a, b}	1 or 3	No criterion for unsatisfactory; just report the score
7. Sampling	4 or 6	Sampling ^a	4 or 6	4 or more
8. Standardized Operations	3	Data Control, Analysis, and Reporting	13	7 or more
9. Managing Data	2			
10. Equating Scores	2			
11. Analyses of Assessment Data	4			
12. Reporting and Using results	2			
Maximum points	26 or 30	Maximum points excluding Linguistic control	25 or 27	
<p>^a Not all countries will answer every question in the Linguistic Control and Sampling groupings. For example, if a country's NLA is a census assessment instead of a sampling assessment, it will have fewer questions to answer.</p> <p>^b Not every country will have a translated version of its NLA. Thus, the questionnaire asks questions about translations but the scoring of this category is not used for determining sufficiency.</p>				

“Sufficient” (good procedural practices):

To attain a “sufficient” score in each of the following four groupings: Capacity and Technical Standards; Instrument Development; Sampling; and Data Control, Analysis, and Reporting
AND

Receive a total (overall) score of 20 or more from the entire 25 or 27 the questions found in the four groupings mentioned above (i.e., excluding Linguistic Control, a country would respond to 25 or 27 questions, depending on whether its NLA was based on a census or sampling, and attain a score of at least 20).