

Expanding Data Collection Tools and Options

1. Background and rationale

The production and dissemination of high quality education statistics are essential for effective planning, as well as for monitoring progress toward national and global education targets. Evidence-based planning reduces system costs by allocating resources more effectively. The added cost of improving data – as long as the data are actually used – is likely much lower than the implicit costs of bad or no information. Planning with bad data inhibits optimal policy implementation, particularly with respect to resource allocation and its equity and efficiency.

Several critical gaps are plaguing the current international monitoring dashboard. Some parts of the education system are not well covered, some populations are excluded and, finally, some aspects of education simply are not measured. In terms of geographical data gaps, Sub-Saharan Africa is the high priority region with low coverage of data. Other data gaps are observed in Eastern Asia and Small Island States. In terms of the data sources, gaps are more frequent for learning assessments and household surveys, whereas administrative data have the highest coverage rate.

The UIS collects administrative and expenditure data on education through the annual [Survey of Formal Education](#). Typically, two-thirds of the 165 targeted and contacted countries respond to the survey every year, although the numbers vary depending on the questionnaire. To produce the international indicators and once national data already mapped to the International Standard Classification of Education (ISCED) is submitted by countries, the UIS uses population estimates from UNPD, GDP from the World Bank, total government expenditure from the IMF, and applies methodologies for indicator calculation as well as protocols for data checking and validation of the produced indicators with the countries, which reduces the number of indicators released.

National institutions can sometimes find challenging to regularly disseminate official statistics on education, especially given the more complex demands of the 2030 Agenda. This is evident in the highly variable nature of country submissions to the UIS but also in the highly-variable quality of official national education information website.

Filling the gaps in current submission is a first step of a broader project, which seeks to improve education data collection, data use, and dissemination. It includes the search and compilation of data that governments have made publicly available through their websites and its use to calculate and report against SDG 4 indicators in compliance with international standards (e.g. ISCED) and methodologies for indicator calculation.

This document discusses the different modalities for data compilation and seeks the support of TCG regarding the combination of modalities of data compilation and harmonization.

2. The UIS Formal Education Survey:

The UIS produces the indicators based on the collection of data, among others, from official administrative sources at the national level through the UIS Education Survey. Collected information encompasses data on educational programmes, access, participation, progression, completion, literacy, educational attainment and human and financial resources. These statistics cover formal education in public and private institutions (early childhood education, primary and secondary schools, colleges, universities and other tertiary education institutions), and special needs education (both in regular and special schools).

The UIS Survey of formal education is based on two main data sources:

- **EMIS systems:** These are usually available from EMIS and used by Ministries of Education for management and planning purposes, and are typically updated on an annual basis. EMIS should typically cover many types of educational paths and levels, including Early Childhood Education (ECE), higher education, and Technical and Vocational Education and Training (TVET).
- **Financial and expenditure data** contain information on different sources of income and expenditure, including government spending on education. This source commonly encompasses data on the construction and maintenance of schools, teacher salaries and household spending on education, including supplies, transport and other costs. Some administrative data are typically available from non-EMIS sources within Ministries but should ideally be linked to EMIS (e.g. school feeding programmes and teacher salaries), depending on how the country organizes the data.

These data are gathered annually by the UIS and its partner agencies through three major surveys¹ being the most important the UIS Survey of Formal Education. The UIS education questionnaires are sent to UNESCO Member States annually. The questionnaires are based on international standards, classifications and measures that are regularly reviewed and modified by the UIS in order to address emerging statistical issues and improve the quality of data.

Questionnaire A: Students and Teachers (ISCED 0 to 4) collects:

- Number of students by level of education, type of institution, intensity of participation, sex, age, and grade
- Number of repeaters for primary and secondary education (ISCED 1 to 3)
- Number of students by programme orientation (ISCED 2 to 4)
- Number of classroom teachers by employment status, qualified and trained status, teaching level of education, type of institution, sex; and annual statutory teacher compensation in public institutions
- Number of educational institutions with ICT services, basic hygiene facilities and the provisioning of life skills–based HIV and sexuality education by level of education and type of institution.

Questionnaire B: Educational Expenditure (ISCED 0 to 8) collects:

¹ The three major UIS Surveys on Education are: UIS Survey of Formal Education, UIS Literacy and Educational Attainment Survey, and UOE Survey of Formal Education. UNESCO-UIS, the OECD and Eurostat (UOE) have jointly administered this annual data collection since 1993. The UOE questionnaire compiles data from high- and middle-income countries that are generally members or partner countries of the OECD or Eurostat. The UOE survey gathers more detailed education statistics.

- Expenditure on educational core services from government, private and international sources (i.e. all expenditure that is directly related to instruction and education, such as teachers, school buildings, teaching materials and books)
- Expenditure on educational services other than instruction (e.g. administration, policy formulation, curriculum development, school feeding and supervision)
- Actual rather than budgeted expenditure
- Expenditure that occurs within the borders of the country (i.e. national schools abroad not included and financial aid to students studying abroad is identified separately).

Questionnaire C: Students and Teachers (ISCED 5 to 8) collects:

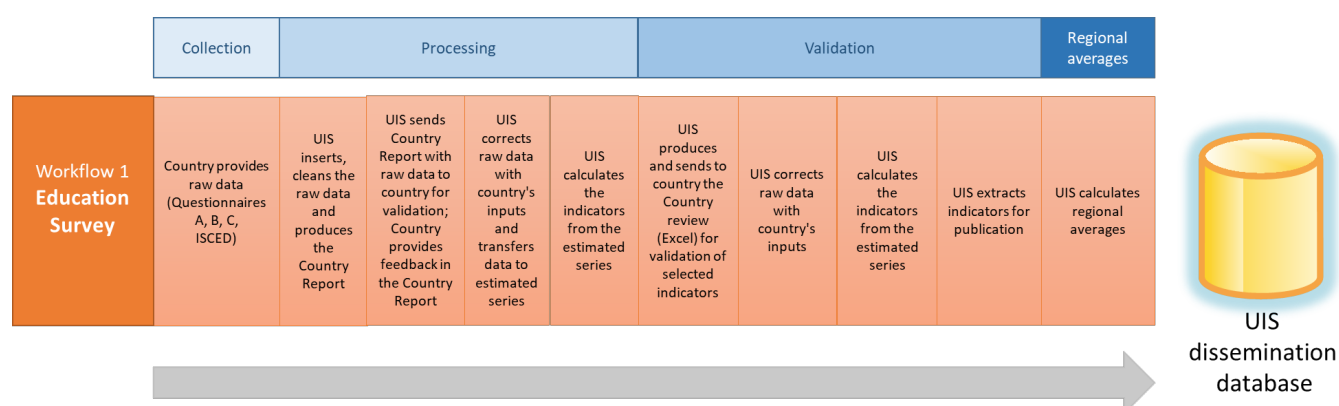
- Number of students by level of education, type of institution, intensity of participation and sex
- Number of new entrants
- Number of graduates by level of education, fields of education, and sex
- Number of internationally mobile students by country of origin and sex
- Number of academic staff by level of education, employment status, type of institution and sex.

Questionnaire on National Education Systems collects:

- Information on national education programmes and their classifications according to the 2011 revision of the ISCED 2011.
- Information used to ensure the production of internationally comparable data and indicators that are gathered through the UIS education questionnaires to produce ISCED mappings, which are essential tools for comparing national education systems.
- Information on the provision of free and compulsory education as per national legislation and/or educational laws.

As shown in Figure 1, the workflow for the production of the indicators takes various steps before the data dissemination, starting from data collection through several questionnaires to many stages of data verification and validation by the member states.

Figure 1: Indicator production workflow



The current process of the Education Survey is fragmented, in the sense that there is a disassociation between the raw data and the indicator figures produced, and it may take several months to complete the cycle:

- **Raw data:** after countries submit raw data, the UIS applies data verification protocols and report inconsistencies (i.e. categories do not add up to total) and/or ask questions (i.e. break in a time series) to countries. Then, when countries reply, which is not always the case, the UIS implements the changes confirmed with countries into the raw data. Sometimes, more than one exchange is needed to finally understand and implement the changes. When there is the peak of data submission, usually around the deadline of the survey, the UIS might delay the report to countries with data inconsistencies or questions for data verification.
- **Indicators:** once the raw data has been verified and before the data release, estimates for indicators are automatically calculated by the UIS and are sent to countries for validation.

In summary, countries submit raw data and get indicator estimates. From one event to the other, it might take several months and exchanges in between.

2.1. Issues associated to current workflow for indicators production

The main issues are:

1. **Heavy demands on the countries to compile** and send data to the UIS and then, months later, to work with the UIS on data checking and validation. These tasks, which are usually executed in addition to other duties by modestly sized units within MOEs, come as a serious burden and are seldom considered as the top priority by already overloaded ministerial statisticians. Due to high staff turnover in some ministries, people who interact with the UIS on data validation are often not the same as those who provided data initially, which may present additional challenges and eventually result in inability to publish certain indicators. Moreover, the situation might repeat with every survey cycle as the personnel responsible for data submission often changes.
2. **Low ownership.** Insufficient knowledge of indicator methodology and the UIS calculation process might contribute to the perception of these indicators as external.
3. **Less than ideal timeliness.** The scope of the UIS operations attempts to unify the process of generation of indicators. Thus raw data are collected, and indicators calculated and validated for hundreds of countries at the same time to be released simultaneously for all. With such set-up, the time lag between compiling data on the country level and releasing statistics by the UIS differs and could be long for some countries. By the moment the indicators are published, the situation on the country level might have changed drastically.
4. **Estimates produced based on low coverage of observations.** This is a big problem for regional averages, which are calculated and published without the full picture of the regions due to data gaps.

3. Dynamic templates, mining and co-building

Aiming to reduce the reporting burden on the countries, the UIS has designed a set of user-friendly tools which help to compile national education data into relevant education levels and categories following international classifications to generate the most relevant indicators.

This complementary workflow could take advantage of publicly available national data that could be mined in templates that have the formulas and metadata embedded and that opens room to discuss and co-create with the countries the indicators.

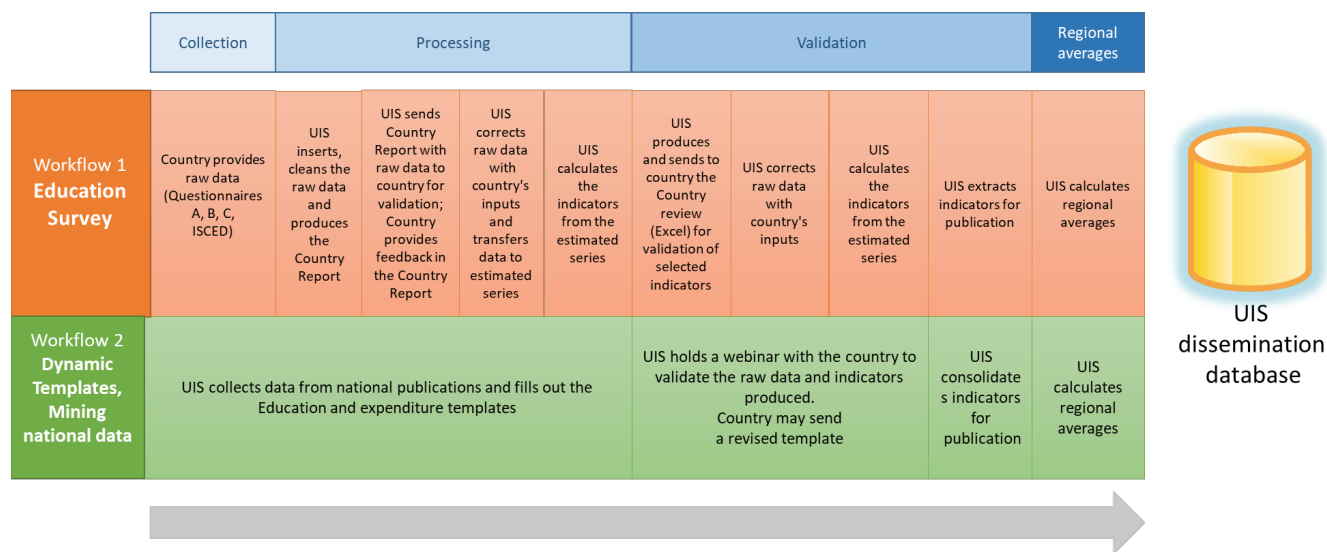
In a nutshell, the strategy is to collect the raw data once national data have been published, to align it to international standards (population, ISCED mapping) and to calculate the indicators in compliance with the methodologies approved by Member States through the [Inter-agency and Expert Group on SDG Indicators \(IAEG-SDGs\)](#) and the Technical and Cooperation Group (TCG).

To deliver on the above a process has been put into place that includes the following steps:

1. Complete the data collection tables, organized by the ISCED levels, with the available information for all the years, including the data and metadata.
2. Verify the resulting indicators with the ones produced by UIS and highlight the outliers for further investigation.
3. Share the results with the countries for the validation and completion of data gaps.
4. Create a consolidated database for all the countries through a routine that auto
5. Create an inventory of the publicly available national publications on education and educational expenditure. Most common sources are Ministry of Education, National Statistical Office and Ministry of Finance.

The workflow (workflow 2 in **Figure 2**) implies that the data collection is started by the UIS by filling in some templates with publicly available national data released by countries, reducing the burden on countries. It also allows a **direct link between the raw data and the indicators estimates** as they are automatically calculated by the template with visible formulas. Therefore, **the verification of raw data and the validation of indicators produced can be done at the same time** through a webinar, workshop or a phone call. This gives more time to countries to focus on the most important activities which are the filling data gaps and validation of the indicators.

Figure 2: Expanding the workflow options










The developed template also helps countries to understand in a very transparent manner how the indicators are calculated following the international methodologies. Moreover, countries can use the template to transfer the national data into international data and indicators without many resources and can use it for their national monitoring and policy discussion. Thus, application on the tools promotes co-creation and development of national capacity, empowering the users and enabling the ownership of the results. These outputs can be included in their own EMIS system in Excel format and it needs to copy and paste in the template.

Finally, through direct interaction with the UIS during completion of the template, the resulting data and indicators can be discussed “on the spot” and, if needed, corrective measures can be applied right away. Thus, validity of the data and indicators obtained through this process can be confirmed by the countries.

3.1. Scope of work: the indicators

The data collection covered 16 SDG indicators: 14 indicators on administrative data and 2 indicators on finance data as shown in the table below:

 Basic education	<p>4.1.3 Gross intake ratio to the last grade</p> <p>4.1.4 Out-of-school rate</p> <p>4.1.5 Percentage of children over-age for grade</p>
 Early childhood education	<p>4.2.2 Participation rate in organized learning (one year before the official primary entry age)</p> <p>4.2.4 Gross early childhood education enrolment ratio in (a) pre-primary education and (b) early childhood educational development</p>
 TVET / Tertiary education	<p>4.3.2 Gross enrolment ratio for tertiary education</p> <p>4.3.3 Participation rate in technical-vocational programmes (15- to 24-year-olds)</p>
 Knowledge	<p>4.7.2 Percentage of schools that provide life skills-based HIV and sexuality education</p>
 Learning environment	<p>4.a.1 Proportion of schools offering basic services, by type of service</p>
 Teachers	<p>4.c.1 Proportion of teachers with the minimum required qualifications</p> <p>4.c.2 Pupil-trained teacher ratio by education level</p> <p>4.c.3 Percentage of teachers qualified according to national standards</p> <p>4.c.4 Pupil-qualified teacher ratio by education level</p> <p>4.c.6 Teacher attrition rate</p>
 Expenditure	<p>1.a.2 Proportion of total government spending on essential services (education)</p> <p>1.a.GDP Government expenditure on education as a percentage of GDP</p>

In the appendix there are some description of the variables and levels collected.

3.1.1. The CESA framework

In order to support the Continental Education Strategy for Africa (CESA), the calculation of the following indicators were added to the template for African region:

- 4.4 & 9.1 Distribution of tertiary graduates by field of education
- 5.2 Percentage of female teachers by teaching level of education
- 5.5 Percentage of tertiary graduates from Science, Technology, Engineering and Mathematics programmes
- 8.1 Distribution of enrolment in secondary education by programme orientation
- 8.2 Percentage of students in technical and vocational programmes

3.2. The dynamic template

As described above, UIS has prepared the package for the countries that include two dynamic templates to collect and calculate SDG indicators for:

1. Administrative data covering the period of 12 years from 2010 to 2021, and
2. Finance data covering the period of 10 years from 2012 to 2021.

Both templates consist of the tables for data collection by year and ISCED level, metadata with definitions, calculation method and data source; and calculation of SDG indicators.

The features of the templates include:

- Automatically produced indicators from the data collection tables;
- For population-based indicators, possibility to select the population data source (national population census or UNPD);
- Possibility to compare the indicators calculated from the national data with UIS published indicator;
- Predefined rules to identify the outliers;
- Automatically produced database for data and metadata.

3.3. *Finding related to the indicators*

The main takeaways are:

- **Disaggregation:** the level of disaggregation of published national data might not meet the level of requirement to produce a given indicator. For instance:
 - 4.1.4 Out-of-school rate: enrolment data by age for pre-primary education or technical/vocational programmes are not available.
 - 4.a.1 Proportion of schools offering basic services, by type of service: data in many cases have not disaggregated by primary, lower and upper secondary education.
 - Data are available only for the national education structure, which is different from ISCED mapping and need additional work.
- **Completeness:** data available for public education institutions only.
- **Definitions:** 4.c.1, 4.c.2, 4.c.3, 4.c.4 for trained and qualified teachers: data compiled from national sources taking into account the national definitions, when available, are different to the one available at the UIS but cannot be compared because the UIS does not collect the national definitions.

3.4. *Lessons learnt*

The main lesson learnt is the evidence that a massive amount of available national data that is not reported to the UIS that can be used to report on SDG indicators.

This situation is more important when the UIS questionnaire, usually responded by the MOE, asks for information that is produced by a different divisions/units/department within a country.

Indicators calculated from national sources need to be validated with the countries as there might be discrepancies between national and international-comparable data.

4. Proposal

Based on the exercise we propose to have a combined indicators production workflow according to countries choice and UIS prioritization:

- a. **Workflow 1- UIS questionnaire:** for the countries reporting using the UIS questionnaire they could preserve this reporting strategy that is already installed and follows a tradition. This is the reporting strategy for all countries reporting to OECD.
 - ✓ **Advantages:** relies on UIS to collect data and produce indicator (i.e.: some recurrent cost to UIS).
 - ✓ **Disadvantages:** subject to country's questionnaire submission.
- b. **Workflow 2 - Dynamic Template, data mining and country's validation:** this is to be agreed with the countries that prefer to engage with this new approach.
 - ✓ **Advantages:** increases government ownership over statistics; allows capacity development on spot; alternatives based on needs (such as use of national population) could be enabled; and values of indicator are validated jointly.
 - ✓ **Disadvantages:** changes as always cause some disruption.

Appendix: Mapping of variables and levels of the dynamic templates

Variables	SDG 4 indicators														Finance indicators	
	4.1.3	4.1.4	4.1.5	4.2.2	4.2.4	4.3.2	4.3.3	4.7.2	4.a.1	4.c.1	4.c.2	4.c.3	4.c.4	4.c.6	1.a.2	% GDP
National education system	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Population data	x	x		x	x	x	x									
Enrolment by sex	x	x	x	x	x	x	x				x		x			
Enrolment by age		x	x	x			x									
Enrolment by grade	x		x													
Repeaters by sex and grade	x															
Teachers by sex										x		x		x		
Trained teachers by sex										x	x					
Qualified teachers by sex												x	x			
Newly recruited teachers by sex														x		
Schools by type of facility								x	x							
Educational expenditure															x	x
Total government expenditure															x	
Gross domestic product (GDP)																x

