Technical Standards

COVID-19: Monitoring Impacts on Learning Outcomes (MILO)

13 April 2021, Version 2.0

The ACER Centre for Global Education Monitoring supports the monitoring of educational outcomes worldwide, holding the view that the systematic and strategic collection of data on education outcomes, and factors related to those outcomes, is required to inform high quality policy aimed at improving educational progress for all learners.

Acknowledgments

Contents

Acronyms

ACER	Australian Council for Educational Research
CONFEMEN	Conference of Ministers of Education of French-Speaking Countries
DM	Data Manager
DMM	Data Management Manual
DTP	Defined Target Population
NC	National Centre
NPM	National Project Manager
NTP	National Target Population
SDGs	Sustainable Development Goals
STF	Student Tracking Form
PASEC	CONFEMEN Programme for the Analysis of Education Systems
PPS	Probability Proportional to Size
QM	Quality Monitor
SC	School Coordinator
TA	Test Administrator
UIS	UNESCO Institute for Statistics

Purpose

1. The purpose of this document is to specify the set of standards to guide the data collection and data management (DM) activities related to assessing the impact of COVID-19 on learning outcomes in six countries in Africa. These standards will enable the creation of a high-quality dataset that allows valid inferences to be made.

2. The COVID-19 Monitoring Impact on Learning Outcomes (MILO) project is a collaborative effort involving numerous stakeholders. It is managed by the UNESCO Institute for Statistics (UIS). The Australian Council for Educational Research (ACER) as a technical partner implements the technical aspects of the project on behalf of the UIS. CONFEMEN as a PASEC coordination partner is responsible for liaising with Francophone participating countries, coordinating the study implementation and providing psychometric data and instrumentation from PASEC 2019.

3. As part of this study, the participating countries will administer:

   • MILO assessment test booklets and questionnaires on Day 1, and

   • National/Regional Assessment test booklets on Day 2.

4. These technical standards relate mostly to the MILO data collection in 2021 as part of the COVID-19 MILO project, but also to the National/Regional Assessment where specified. The technical standards can also be adapted and used to guide future data collection cycles.

5. The standards for data collection and submission are developed with three major and inter-related goals in mind: consistency, precision and generalisability of the data. Furthermore, the standards serve to ensure a timely progression of the project in general.

   a) Consistency: Data should be collected in an equivalent fashion in all schools, using equivalent test materials. Given consistent data collection (and sufficiently high sample sizes and response rates), test results are comparable across sub-populations. The test results of different sub-populations will reflect differences in the performance of the students measured and will not be caused by factors which are unrelated to performance.

   b) Precision: Data collection and submission practices should leave as little room as possible for spurious variation or error. This holds for both systematic and random error sources, e.g., when the testing environment differs from one group of students to another, or when coding procedures leave room for interpretation. An increase in precision relates directly to the quality of results one can expect: The more precise the data, the more

powerful the (statistical) analyses, and the more trustworthy the results to be obtained.

c) Generalisability: Data are collected from specific individuals, in a specific situation, and at a certain point in time. Individuals to be tested, test materials and tasks etc. should be selected in a way that will ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected, but hold for a variety of settings and are valid in the target population at large. Thus, collecting data from a representative sample of the population, for example, will lead to results that accurately reflect the levels of numeracy and literacy of students at the end of primary school education (SDG 4.1.1.b Minimum Proficiency Levels in reading and mathematics) in the defined target population.

d) Timeliness: The complexity of the tasks of this study makes it necessary for UIS, ACER, CONFEMEN and participating countries to adhere to pre-determined timelines and facilitate timely turnaround of communications and tasks. Therefore, general standards relating to timelines and communication will be included. Task specific timelines will be incorporated where it is practical to do so.

6. There are two types of standards in this document; each with a specific purpose:

- Data quality standards refer to aspects of study implementation that directly concern the quality of the data or the assurance of that quality.

- Project management standards are in place to ensure that all assessment operational objectives are met in a timely and coordinated manner.

Data Quality Standards

Target population and sampling

Rationale

The following population definition and sampling standards aim to achieve a level of precision in line with the best practices from established large scale assessment surveys[1].

Meeting the standards specified in this section will ensure that the assessed students come from the same target population. To be able to draw conclusions that are valid for the entire population, a representative sample shall be selected for participation in the assessment for students at the end of primary education. The representative sample should be large enough to achieve a desired precision of measurement. For this reason, minimum numbers of participating students and schools are specified.

Standard 1.1

The UIS, ACER, CONFEMEN and the participating countries will work together to identify the population to which inferences will be made. This population is referred to as the Desired Target Population.

The target population for the study is all students in the final grade of primary school where the language of instruction is English or French. By agreement, countries may use a grade other than the final grade of primary school where that is the grade that a country references to report against SDG indicator 4.1.1(b).

All students enrolled in the final year of primary school in recognised schools where the language of instruction is either English or French belong to the National Target Population (NTP). In simple terms, the NTP is intended to provide full coverage of all eligible students in the education systems of participating countries of Africa. Any deviation from the full national coverage must be described and quantified in advance.

Standard 1.2

The Defined Target Population (DTP) covers 95% or more of the NTP. The UIS, ACER, the CONFEMEN and the National Centres (NCs) will work together to identify schools and students that are impractical to assess. These exclusions are referred to as school-level exclusions and within-school exclusions. The total of the combined school-level exclusions and student level exclusions will be no greater than 5% of the NTP.

Standard 1.3

Only students within the DTP participate in the MILO assessment.

---

[1] Such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA).

Standard 1.4

The school sample will be drawn using established and professionally recognised principles of scientific sampling.

The sampling design for the MILO assessment is a two-stage stratified sample design. The first-stage sampling units consist of individual schools having students in the grade corresponding to the grade corresponding to the final year of primary school. Schools are sampled systematically from a school sampling frame, with probabilities that are proportional to a measure of size. The measure of size is a function of the estimated number of the MILO assessment-eligible students enrolled in the school for the corresponding grade. This is referred to as systematic Probability Proportional to Size (PPS) sampling. Sampling procedures are based on these principles.

Standard 1.5

The second-stage sampling units consist of students belonging to the schools selected in the first-stage sampling. The second stage of sampling is conducted in cases where the number of students in selected schools belonging to the target populations exceeds the target cluster size. In all other cases, all students in the selected schools that belong to the target populations are selected in the sample.

A sampling tool or method specified by ACER will be used to scientifically draw random samples from lists of students in each sampled school.

Standard 1.6

ACER will work with the key stakeholders to set the sample size to achieve a level of precision in the sample estimates for each country equivalent to a 95% confidence interval of 5 percentage points for estimates of percentages, or 0.1 of the population standard deviation in student achievement for estimates of mean scores.

Standard 1.7

The school sample size needs to result in a minimum of 150 participating schools[2]. For each sampled school, two substitute schools will be selected where possible, using a systematic method, to ensure a proper level of school response as indicated in Standard 1.9.

Standard 1.8

The student sample size is a minimum of 4,500 assessed students.

---

[2] To simultaneously achieve Standard 1.6, it is likely that at least 250 schools will be required.

Standard 1.9

The school response rate must be at least 85% of sampled schools. If the response rate from sampled schools does not reach this level, then substitute schools may be used to reach an acceptable response rate.

Standard 1.10

The student response rate is at least 80% of all sampled students across responding schools. This response rate includes students from substitute schools.

Standard 1.11

Absent sampled students cannot be substituted with non-sampled students.

Standard 1.12

Sample weights will be calculated to reflect the contribution of each participating student to the survey estimates, taking into account the sample design and adjustments for non-response.

Note 1.1   A student is regarded as a participant if they have responded to 10% or more of the assessment items

Note 1.2   Data from schools where the student response rate is greater than 25% will be included in the dataset

Note 1.3   For the purposes of calculating school response rates, a participating school is defined as a sampled school in which more than 50% of sampled students respond

Note 1.4   School level exclusions are schools which may be excluded from the sampling frame because:
- of geographical inaccessibility
- of extremely small size (<5 eligible students)
- all students within the school would be within-school exclusions
- of other agreed reasons

Note 1.5   Student level exclusions are exclusions of particular students from the assessment because of one of the following:
- the student has a functional disability – that is, physical disabilities such that they cannot perform in the assessment situation.
- the student has an intellectual disability – that is, students who, in the professional opinion of the school principal or other qualified staff, are emotionally or mentally unable to follow the general instructions of the assessment.
- the student has insufficient language experience – that is, students who are unable to read or speak the language(s) of the assessment and would not be able to overcome the language barrier. Such students meet all of the following criteria:
* they are not native speakers of the assessment language
* they have limited proficiency in the assessment language
* they have less than one year of instruction in the assessment language

Language of testing

Rationale

Learning outcomes are most accurately reported for those students for which the test is administered in a familiar language. In instances where students have limited knowledge of the testing language, the test may underestimate their ability. It is therefore optimum that the test is administered in the language of instruction.

Standard 2.1

Test booklets and questionnaires will be administered in either English or French. The NC will determine which is the appropriate language of assessment for their respective education systems.

Test development

Rationale

Instruments should reflect the requirements outlined in the MILO Assessment Blueprint and Contextual Framework. The instrument should provide data that can be analysed to address the research questions of the MILO assessment. The tests should provide fair and accurate measures of students' achievement on the domain which is defined by the blueprint and they should adhere to the test specifications. The questionnaires should address issues which are specified in the research questions of the MILO assessment to maximise reliability and validity of the measures and to minimise the burden on respondents.

Standard 3.1

An assessment blueprint and a contextual framework will be developed detailing the specifications for the test booklets and questionnaires. These documents will describe the content of the instruments, the way that they will cover the different constructs of the domain, types of items, the timing and the conditions under which the instrument is administered.

Standard 3.2

The test booklets meet the domain definitions and test specifications. All aspects of the test are clear and accurate including the stimulus, items and instructions. The items are appropriate to the purpose of the test, the population and the test specifications.

Standard 3.3

The procedures used to develop the test booklets are clearly documented, including detailing the quality assurance processes that are used.

Adaptation/ translation and linguistic verification of material

Rationale

In order to ensure that measures derived from assessment instruments are comparable within the country it is necessary to use a set of standardised items. Efforts should be made to ensure that each adapted item and booklet are relevant to the target population and equivalent to the source version. Specific terms within MILO contextual questionnaires need to be adapted in such a way to ensure their comparability. A lack of adaptations or inappropriate adaptations can jeopardise the comparability of data.

Similarly, it is essential that equivalent information is provided to all students participating in the assessment. Any instructions given to the students, as well as the procedures used throughout the test administration need to be equivalent. To achieve this goal, all individuals who play a key role in the data collection process, i.e., the Test Administrators (TAs) and School Coordinators (SCs), should receive and deliver a set of standardised instructions.

Standard 4.1

Both English and French versions of MILO student test booklets and contextual questionnaires are conceptually equivalent. Agreed upon MILO questionnaire adaptations to the language-specific context are made if needed.

Standard 4.2

The following documents are translated into the French language in order to be linguistically equivalent to the English source versions.

- MILO test booklets
- MILO contextual questionnaires (student, school and system level questionnaires)
- The MILO TA script from the TA manual.

Standard 4.3

The English source version of MILO testlests and contextual questionnaires translated into French will be independently verified prior to implementation to ensure that generic and item-specific translation guidelines have been followed.

Printing of material

Rationale

Variations in print quality may affect data quality. When the quality of paper and print is poor, the performance of students may be influenced not only by their levels of proficiency, but also by the degree to which test booklets and contextual questionnaires are legible. To rule out this potential source of error, and to increase the consistency and precision of the data collection, paper and print quality samples are required from the NC.

Standard 5.1

All student test booklets and the contextual questionnaires are printed using high quality paper and print quality. They will be printed on 80gsm (grams per square metre) paper.

Standard 5.2

The cover page of both MILO and national/regional test booklets and MILO contextual questionnaires used in schools contains all information as specified by ACER and outlined in the NPM manual.

Standard 5.3

The format, pagination and layout of both English and French language versions of the MILO test booklets are equivalent.

Standard 5.4

The format, pagination and layout of both English and French language versions of the MILO contextual questionnaires are equivalent.

Test administration

Rationale

Certain variations in assessment procedures are likely to affect test performance, such as the session timing, the administration of test materials and instructions given prior to and during testing, and rules for excluding students from the assessment. A full list of relevant MILO assessment conditions is given in the assessment operational manuals.

The TA plays a central role in the assessment procedures. Special consideration is therefore given to the training of the TAs, ensuring that as little variation in the data as possible is caused by random or systematic variation in the activities of TAs.

The MILO assessment covers a wide range of content areas. Given the time constraints, the test booklets include clusters of test items on a rotated basis, and test booklets are allocated to students in a statistically random fashion. Student Tracking Forms (STFs) will be used to ensure the correct allocation of MILO test booklets to students by TAs.

Standard 6.1

All MILO assessment sessions follow the procedures as specified in the TA manual, particularly the procedures that relate to:

- Test session timing
- Student tracking:
    - a STF is prepared for each sampled school administrating the MILO and the National/Regional Assessments
    - MILO test booklets are distributed to students according to the order specified in the STF
- Maintaining the MILO assessment conditions.

Standard 6.2

TAs are trained in the field operations procedures outlined in the TA manual. TAs receive a copy of the TA manual prior to the MILO assessment session.

Standard 6.3

TAs read out the standard TA script prior to the students sitting the MILO assessment session.

Standard 6.4

The relationship between TAs and participating students must not compromise the credibility of the MILO assessment session. The TA will be independent from the students and the school staff, which means he or she should **not** be:

- An instructor of any student in the MILO assessment session he or she will administer
- A member of staff in the school in which he or she will administer the MILO assessment
- A relative of any of the staff in the school in which they administer the MILO assessment
- A parent or close relative of any of the participating students in the MILO assessment.

Security of material

Rationale

The goal of the assessment is to improve the quality of education through measuring students' learning outcomes and understanding the contextual factors associated with learning outcomes in order to estimate the impact of COVID-19 on learning and to evaluate the effectiveness of distance learning mechanisms utilised during school closures. Prior familiarisation with the assessment materials, or training of students to the assessment, may affect the validity and comparability of the data, and changes in learning outcomes. Therefore, confidentiality of the assessment materials is essential.

Standard 7.1

The MILO and National/Regional Assessment materials designated as secure are kept confidential at all times. Secure materials include all test materials, data and draft materials. In particular:

- Only approved project staff and participating students during the test session are able to access and view the test booklets.

- Only approved project staff have access to secure data and embargoed materials.

Standard 7.2

Formal confidentiality arrangements are in place for all approved MILO project staff.

Quality monitoring

Rationale

To obtain valid results from the assessment, the data have to be collected in a consistent, reliable and valid fashion. Independent Quality Monitors (QMs - observers) are responsible for assessing the implementation of activities that align with this goal during the test administration.

Standard 8.1

The MILO test administration is monitored using school visits by trained independent QMs.

Standard 8.2

At least 5% school visits are conducted in each participating country to observe MILO test administration sessions. A range of different types of schools will be included in the sample for monitoring.

Standard 8.3

MILO Test administration sessions that are the subject of the national QM visit are randomly selected.

Standard 8.4

QMs will be familiar with the test implementation procedures of the MILO assessment, complete the quality monitoring checklist and observation form, and be familiar with the education system of that country. Also, QMs must not have a personal interest in the results of the school or be personally affiliated with the observed school.

Data management

Rationale

To obtain valid results from the assessment, the data collected must be of a high quality, using consistent, reliable, and valid approaches. Consolidating and merging national databases is a time-consuming and difficult task. To ensure the timely and efficient progress of the project, ACER needs continuous access to national staff helping to rule out uncertainties and to resolve discrepancies. This standard aims to prevent substantial delays to the whole project which could result from a delay in processing the data from one or more NCs and to avoid the loss of the data.

Standard 9.1

Each NC should appoint a data manager (DM). Data managers will be required to attend training on data management and use of ACER Maple data management software, provided by ACER prior to data collection. Data managers will train NC data administration and data entry personnel on the use of ACER Maple in line with the procedures described in the Data Management Manual (DMM).

Standard 9.2

ACER Maple software is used for data management, data entry and preliminary data verification prior to submitting data to ACER.

Standard 9.3

The data verification procedures, as specified in the DMM, will be executed by the NC staff in ACER Maple software before submitting the final database to ACER.

Standard 9.4

A data manager from the NC will be available upon submission of the database to ACER. The data manager:

- is authorised to respond to ACER data queries

- is available for a four-week period immediately after the database is submitted unless otherwise agreed upon

- is able to respond to ACER queries in English within three working days

- is able to resolve data discrepancies.

Data submission

Rationale

The timely progression of the project depends on the quick and efficient submission of all collected data. Therefore, participating counties are asked to submit only one standard database to ACER.

Standard 10.1

Participants' tracking data and test booklets data for both MILO and National/Regional Assessments, and MILO contextual questionnaire data collected by NC must be entered into ACER Maple data management software provided by ACER, as specified in the DMM.

Standard 10.2

Each NC submits its data to ACER in a single database in the specific format produced by ACER Maple software, as specified in the DMM.

Standard 10.3

NC submits its data to ACER after the data for all instruments and all participants are entered into ACER Maple and all discrepancies are resolved, as described in the DMM.

Standard 10.4

All data are submitted without recoding any of the original response variables.

Standard 10.5

The timeline for submission of national databases to ACER is within eight weeks of the last day of assessment, unless otherwise agreed between the NC and ACER.

Psychometrics and data analysis

Rationale

The production of a high-quality database and the use of modern psychometric and statistical methods is essential to the integrity of the MILO and the National/Regional Assessments. A high-quality database will ensure that researchers can analyse the data in a standard way, following methods established in high-quality large-scale education surveys. Following standardised procedures will help ensure that the MILO and the National/Regional Assessments databases are consistent and comparable.

Standard 11.1

Data will be cleaned. Anomalies regarding duplicate identification variables, out of expected range values, and invalid codes will be resolved and reported.

Standard 11.2

Sample weights will be calculated and included in the final database.

Standard 11.3

For assessment data, missing responses are scored incorrect, and all trailing missing are treated as not administered except for the first in the sequence, which is treated as incorrect for the item calibration stage.

Standard 11.4

Assessment data will be scaled using models derived from Item Response Theory (IRT). The choice of model will be approved by the UIS.

Standard 11.5

Differential Item Functioning (DIF) by gender and by language will be calculated and reported. Treatment of items showing DIF will be reported.

Standard 11.6

Item statistics will be reported. These will include indications of missing, facility, item-rest correlations, estimates of ability across category and estimates of goodness of fit to model.

Standard 11.7

Learner ability and item difficulty estimates will be placed on a single scale covering the end of primary school for each of reading and mathematics.

Standard 11.8

A Plausible Values method will be used to generate unbiased population estimates of learner ability. Plausible Values will be included in the final MS database.

Standard 11.9

Conditioning of the psychometric population model will be used to improve sub-population estimates. Conditioning on gender and participating country, at least, will be implemented.

Standard 11.10

Sampling variance will be calculated using an appropriate method, such as sample replication or linearization. The choice of method will be documented. If replication is used, replicate weights will be included in the final MS database.

Standard 11.11

All analysis of assessment data will include adopting Plausible Value methods, using sample weights, and using appropriate methods for determining sampling variance. Standard Errors for all statistics will be provided to the UIS.

Project Management Standards

Communication

Rationale

To ensure the timely progression of the project, delays in communication among all parties involved should be minimised.

Standard 12.1

Qualified staff from all parties involved in the processes of the data management and data submission are available to communicate and respond to queries during all stages of the project based on the work plan activities.

Schedule for submission of materials

Rationale

In order to progress according to project timelines, efforts should be made to ensure that all parties involved can submit project materials within the allocated timeframes.

Standard 13.1

All parties involved will keep to pre-determined schedules for all activities, including:

- population definition and stratification variables
- assessment window definition
- sampling
- contextual questionnaire item review
- translation review
- contextual questionnaire adaptation
- data submission
- reporting.

Archiving of materials

Rationale

The NC will maintain an archive of electronic and paper forms of all assessment material for both MILO and National/Regional Assessments. This will provide an overview of all materials. This will also ensure that instruments will be available to all parties involved to assist with data cleaning and processing.

Standard 14.1

The NC will maintain a permanent electronic archive of all assessment materials for both MILO and National/Regional Assessments, including:

- all administered test booklets
- all administered contextual questionnaires
- sampling forms and sampling frame
- school sample results and selection numbers
- tracking forms
- QM assessment forms.

Standard 14.2

The NC will archive and retain all assessment materials for both MILO and National/Regional Assessments a minimum of three years. Materials to be archived include:

- all completed test booklets and contextual questionnaires in paper format
- student lists
- student tracking instruments
- all submitted data.

Standard 14.3

Archived materials will be stored in one location only and will be under the guardiancy of the NC.

Standard 14.4

Upon request by the appropriate authority, materials will be deleted from all electronic sources and physical materials shredded in line with the requirements of national laws.

Reporting and dissemination

Rationale

The success of the study is dependent upon the reporting and sharing of results with all relevant national stakeholders. This reporting includes the conclusions derived from national data collection. It is important to ensure that the circulation of information occurs within time periods that are appropriate to both key stakeholders and the project as a whole.

Standard 15.1

The NC develops appropriate mechanisms in order to promote participation and effective implementation.

Standard 15.2

Adherence to the Technical Standards will be monitored and reported by the NC to the UIS.

Standard 15.3

The reporting will indicate the relative levels of achievement of students in each of the two subjects.

Standard 15.4

The reporting will include the findings regarding the relationship between background variables and learning outcomes.

Standard 15.5

The reporting will include contextual findings in their own right relevant to the study research questions in relation to the students, schools and systems.