WG/GAML/14

LEARNING THROUGH THE LIFECYCLE

IN MULTI-TOPIC NATIONAL HOUSEHOLD

Note: the same document is also reviewed by the Working Group on Household Survey (ref.: WG/HHS/8)

I. Summary, Conclusions and Next Steps of the meeting on 3-4 February, 2020 in Washington, DC

Background

On February 3-4, 2020, experts on learning assessments and from major international survey programs gathered at the World Bank's Headquarters in Washington DC to discuss challenges and opportunities to address gaps in the availability of learning data in low- and middle-income countries. The workshop, entitled *"Learning through the Lifecycle in Multi-Topic National Household Surveys: Options for low-income countries"*, was convened by the Learning Assessment Platform (LeAP) team, the Skills Thematic Group in the Education Global Practice, and the Living Standards Measurement Study (LSMS) team in the Development Data Group at the World Bank and UNESCO Institute for Statistics (UIS). The meeting brought together a mix of experts and institutions active in the collection of learning and skills assessment via different modalities such as, besides the World Bank and UNESCO UIS, UNICEF, World Bank, the Demographic and Health Surveys (DHS), OECD, Westat, ETS, national agencies and citizen-led assessment organizations from Mexico, Tanzania, and the West Bank.

The focus of this workshop was on the challenges in reporting on the Sustainable Development Goals (SDGs), the importance of generating learning data, and the lessons learned from international experiences in this area. The objective of the workshop was promoting the harmonization of survey instruments for the assessment of learning across national and international survey programs, including: (i) practical guidance for addressing the technical challenges, (ii) identification of the knowledge gaps that need to be closed, and (iii) development of a network of experts and institutions that can work together to identify and support opportunities for skills and learning data collection in household surveys.

The workshop agenda was structured in three parts. The first two parts were a stock-taking of the discussions around the learning assessment modules for pre-school (ECD), school-age children and adults with a focus on the national, regional and international programs. Then, lessons learned from different implementation experiences guided the discussion into exploring examples of technical issues (e.g. sampling strategies, CAPI versus PAPI tradeoff administration) that need to be resolved to facilitate data collection of learning and skills assessment in household surveys. The final session was dedicated to charting a practical way forward in terms of both content and process.

This note summarizes the key points discussed over the two days, and identifies some practical follow-up actions aimed at furthering the agenda towards an internationally coordinated initiative to address gaps in learning data through the lifecycle. More details about the workshop are also being made available by the organizers on the web.

Summary of the discussions

Introduction. The workshop commenced with a discussion on current practices and the way forward, and the need to generate more learning data to fill current gaps in low- and middle-income countries. It was noted that a learning assessment can only be universal when it reaches all children (even those not in school), and that household surveys offer much better scope for addressing issues around inequities in learning compared to school-based assessments. The scope of the data gap is daunting, pointing to the importance of finding synergies among international partners and countries working on learning

data collection in the monitoring of SDG 4. The outcome of an international effort should serve global monitoring goals, while also being providing the information needed by national level decision makers.

Session One: This session focused on the instruments and implementation experiences for measuring children's skills and learning. There is limited availability of comparable measures of children's developmental outcomes at the population level. Current developments aimed at filling that gap include the methodological development of an updated ***Early Childhood Development Index (ECDI)*** measurement tool and its inclusion in Multiple Indicator Cluster Surveys (MICS), as well as some DHS surveys, to track SDG 2.4. The ECDI has been validated through a rigorous and consultative process and has been presented to the UN Statistical Commission on various occasions.

Another tool discussed was the ***Foundational Learning Skills (FSL)*** module in household surveys and its use around the world. Applied to children age 7-14, the FSL includes early reading and early numeracy tasks aligned with grade 2 and grade 3, and this is relevant because a large proportion of teenagers globally are unable to understand grade 2 level texts. This concern makes the study of variances in learning assessment scores and the associated factors even more paramount.

The implementation of learning/skills modules in household surveys was also discussed. The OECD representative discussed the development and use of ***PISA for development (PISA-D)*** to monitor SDG 4.1.1c and the experience of surveying out-of-school participants in the countries taking part in PISA-D. The main challenge here is that there are high screening costs involved in the effort to identify out-of-school children in low and middle-income countries. The INSP team from Mexico showed the country's experience with the ECDI tool as used in their national household survey in 2018. The main challenges were encountered in training of enumerators, data quality control, translation of the tool, reaching respondents, and evidence of security issues in specific locations. The team from UWEZO, Tanzania also discussed the role of citizen-led assessment efforts in Tanzania to measure foundational literacy and numeracy skills of out-of-school children.

Session Two: The second session focused on the instruments and the implementation experiences for measuring adult skills and learning. To this end, UNESCO UIS presented several tools used to monitor SDG 4 that could be included in household surveys such as the ***mini-LAMP***, which is a tool used to assess basic literacy and numeracy skills with a small number of items. The discussion emphasized the relevance of developing tools that can be administered in a short period of time and can be available in several languages, depending on the level of proficiency that the country is interested in.

During a panel on the implementation of learning modules for adults in household surveys, the UNESCO GEMR team presented results of the digital (ICT) skills module currently included in MICS, which allows for disaggregation by age, gender, wealth and location. This digital skills module is also highly correlated with PIAAC. The experience of the LAMP pilot in Palestine faced significant implementation challenges during the assessment administration, such as difficulty to find suitable respondents to apply the Reading component module due to the high literacy rate in Palestine, long-time lag between the pilot test and the main interviews (over 5 years), etc. Standardized tools can meet the needs of countries, but the countries may need some flexibility when adopting them for population learning monitoring.

Session Three: The final session focused on design choices to address the gaps in cognitive skills data in low-income countries. The first topic of this session was on survey design, and the LSMS team at the World Bank presented the main design options (i.e. whether to use the full or abbreviated version,

conduct learning assessments for all or a sub-sample of households interviewed in multi-topic surveys, etc.) and related administration issues (i.e. sampling selection of households and household members, mode, time and length of the interview, enumerators skills, frequency, etc.) for integrating learning assessments into multi-topic surveys. When integrating a stand-alone assessment into household surveys, it is important to consider factors such as additional costs, timing, number of visits, and self-reporters versus proxy respondents.

The second topic of the session was on **sampling design** and representatives from Westat presented some examples of sampling designs from applications of PISA for Development (PISA-D) in low- and middle-income countries. PISA-D is the first PISA survey to collect data outside the school (i.e. in households or similar locations) and develop robust sampling procedures by taking into account the varied demographic profiles of PISA countries and economies, and ensuring each sample adequately represents the target population of all 14 to 16 year-olds who are out of school, or in school but in grades 6 or below. Censuses collect age and school enrollment information of the adult population that can be used as a frame for an estimated measure of the population. Household surveys could also be used to select the population for the learning module. This would involve oversampling the population group of interest, adding screening questions to the main household surveys and collecting information on non-respondents.

The third topic of the session was on **survey administration tools, PAPI versus CAPI**. The main advantages of most of the CAPI applications are: (i) the creation of complete log files, recording all actions carried out during the interview (i.e. keystrokes, response time, back-ups, etc.), (ii) the validation and test for outliers or unusual values, which allows for the immediate detection and correction of errors, and (iii) the correction of an imperfect sampling frame by using digital maps and ensuring sufficient coverage in the second phase. There are methodological and logistical advantages of using CAPI, such as durability, functionality, and savings in printing and delivering. Potential disadvantages include the additional costs of programming, problems with servers and data storage off-line, and lack of technical know-how.

The fourth topic of the session was on **adaptive testing trade-offs**. The session covered pros and cons of adaptive testing as well as its use in some large scale programs. Adaptive tests allow tailoring a test for the ability level of examinees. Adaptive algorithms optimize the delivery of tests items to match the characteristics of individuals, thereby allowing the test to provide more reliable information about skills in a relatively shorter time. Depending on the design, these algorithms can tailor both the difficulty and number of items to the characteristics of a respondent, thus maximizing precision of measurement across a larger range of abilities. While adaptive tests provide an opportunity to accumulate greater accuracy in making the decision, they require more sophisticated methodologies for analysis and larger sample sizes. Furthermore, adaptive algorithms assume that automatically scoreable items are used, so that items that cannot be automatically scored are not usable. The benefits of adaptive testing include (i) Improved measurement without increasing testing time; (ii) At the individual level, benefits in terms of time and efficiency of measurement by ensuring students answers items around their levels of ability, and (iii) At the country level, it better targets the average ability of test-takers.

The last topic of the session was on **localization issues**, based on the experience with the implementation of the MICS Foundational Learning Skills Module in various countries. Issues covered included customization to context and cultural relevance, choice of language(s), adjusting the age range, ethical considerations, privacy and protection, methodologies for enumerators training and more. It is important to fully document the localization of instruments to facilitate data harmonization and the

scale-up of measurement efforts. Regarding the language for the assessment, different combinations of analysis should be explored based on the language of instruction and the language spoken at home.

Conclusions and Next Steps

There was substantial extent of agreement among the participants that this is an opportune moment for scaling up the collection of learning data, and that the link to international survey programs is in most instances an appropriate one to ensure progress is made at scale and in a cost effective and sustainable manner. There is also a good recognition, however, that there are real implementation challenges and that a substantial amount of coordination and effort will be needed across agencies to realize the vision of closing the data gap in learning data in low- and middle-income countries. There was also a clear recognition that that there may be different design options that would permit leveraging those linkages and that good protocols and standards in terms of implementation are needed.

The discussion at the workshop have shown that recent advances made in developing learning assessment tools for different age groups offer a suite of tools to address the information needs related to the different age groups: ECDI for early childhood, FLS for 7-14 years old, PISA-D for 14-16 years old, and mini-LAMP for adults. With some variation, these tools have been tested and validated, and come in reasonably sized modules that make them amenable to an implementation model coordinated with that of a household survey.

The coalition of partners present at the workshop has therefore agreed to engage in developing a coordinate strategy to move from aspiration to implementation. The following practical next steps were agreed.

- Preparation of a stock-taking document on available learning assessment tools, including advantages and disadvantages of each instrument (Timeline: end of March; Lead: UNESCO UIS);

- Establish a task force comprised of education and survey specialists to develop a concrete roadmap for improving learning data collection in household surveys (Timeline: end of March; WB to coordinate; UNICEF, UNESCO-UIS to nominate their members);

- Task force to develop a Guidelines document. The Guidance document will be guided by the objective of developing a tool or set of tools to capture individual abilities to meet certain minimum thresholds ('foundational skills'), and include specific, prescriptive, practical proposals to resolve some of the critical implementation trade-offs that emerged at the workshop (Timeline: End May);

- Identify 2-3 countries to pilot the approach(es), and develop a strategy to jointly fundraise to get implementation off the ground and for the eventual scale-up;

- Participation in the OECD Meeting in Paris (first week of June) for discussing the guidance document

II. Terms of Reference for review of learning assessments in household surveys

Background

While countries have committed to report on the multiple education-related indicators under Sustainable Development Goal (SDG) 4, much of this effort has focused on SDG 4.1.1, which measures minimum proficiency in reading and mathematics at different grade levels, and which uses large-scale assessments (either international or national) for data collection. Other SDG indicators, such as 4.5.1, which measures gender, rural/urban, wealth and other parity in education outcomes, and 4.6.1, which focuses on the literacy and numeracy skills of youth and adults, would require data collection to happen outside of the classroom. Multi-topic household surveys that are regularly carried out in most developing countries around the world offer a possible vehicle to collect such data at scale.

There are several reasons why international or national large-scale assessments of in-school students could be complemented by assessments of skills or student learning via household surveys. Such household surveys can allow for capturing a representative sample of out-of-school population, be it due to age (i.e. pre-school or post-school) or due to education status (i.e. school-age dropouts). Likewise, the ability to correlate assessed skills or competencies with a diverse set of individual-, household-, and community-level variables collected as part of these surveys can provide significant scope for better understanding the factors that could be driving learning outcomes. Furthermore, such surveys allow for triangulating direct assessment of an individual's skills or competences with other-reports by household members (e.g., direct assessment of a preschooler's skills and caregiver report on these skills) to derive a more comprehensive assessment of these skills and competences. Finally, collecting learning assessment data via household surveys allows for the linkage of data on learning to administrative and service delivery data (on health, education, transport, etc.).

Moreover, given school closures in the context of COVID-19, system-level assessments of student learning, which are usually administered in classrooms, have been put on hold. Given the uncertain timeline for the return to in-person schooling, integration of learning assessment into one-on-one in person or phone-based survey administration can address the need for stakeholders to be informed of students' learning status and to measure the learning losses in a systematic and rigorous manner.

In this context, the World Bank, UNESCO Institute of Statistics, UNICEF, and OECD are collaborating on a work program that would enable countries to administer short learning assessments as part of household surveys. A February 2020 technical workshop, "Measuring Learning through the Lifecycle in Multi-Topic National Household Surveys: Options for low-income countries," brought together experts on learning assessments and specialists from major international organizations, and discussed recent advances made in developing short learning assessment tools for different age groups that could be implemented as part of a household survey.[1]

The objective of this consultancy is to prepare a document that takes stock of pre-identified learning assessment tools developed to capture individual abilities to meet certain minimum thresholds ('foundational skills'), which have already been (or are ready to be) implemented as part of a household survey. The document will include information on evidence of validity and reliability of these tools, and lessons learned from their administration in different settings. Likewise, the document will highlight the

---

[1] These include the UNICEF Early Childhood Development Index (ECDI) for early childhood, UNICEF Foundational Learning Skills (FLS) for 7-14 years old, OECD PISA for development (PISA-D) for 14-16 years old, and UIS mini-LAMP (Literacy Assessment and Monitoring Programme) for adults.

potential of the existing tools to be administered through different delivery modalities (one-on-one in person interview, self-assessment, and phone-based) in the future.

Deliverables, timeline and payment schedule

The consultant will be expected to complete the following two tasks:

1) Provide a summary of the key features of assessment instruments – namely **UNICEF Early Childhood Development Index (ECDI) for early childhood, UNICEF Foundational Learning Skills (FLS) for 7-14 years old, OECD PISA for development (PISA-D) for 14-16 years old, and UIS mini-LAMP (Literacy Assessment and Monitoring Programme) for adults.**[2] Key features should include:

   a. Subjects / skills assessed

   b. Age groups assessed

   c. Administration duration of the module

   d. Inclusion of adaptive testing

   e. Administration method (direct assessment, caregiver report, etc.)

   f. Collect all items of the respective assessments which could be used in traditional household survey

2) Summarize experiences of fielding these learning assessment instruments, by creating a database that, for each instance of fielding of an instrument, reports on:

   a. Country of assessment

   b. Language(s) of assessment

   c. Sampling strategy and sample size

   d. Administration modes (paper and pencil, computer-based, tablet-based, phone-based, other)

   e. Stand-alone or with other modules (background questionnaire, socioemotional skills /mental health, household survey modules, etc.)

   f. Validity and reliability statistics of learning assessment

   g. Important correlations with other modules, if any

   h. Implementation challenges / lessons learned

3) Conduct a desk review of phone-based[3] learning assessments that are used to monitor learning at the system level, including key feasibility conditions (including type of item, hardware needed for both the interviewer and the respondent, financial and human resource inputs needed), evidence of validity and reliability of each assessment, and key challenges and lessons learned in implementing such assessments.

---

2 Any additional instruments to be reviewed will be discussed by the team and the consultant.

3 While the general scope of the assignment focuses on delivering "phone-based" assessments via computer-assisted telephone interviewing (CATI); the desk review should also include any relevant experience conducting learning assessments for system-wide monitoring via interactive voice response (IVR) and SMS technologies.

4) Assess the feasibility of phone-based administration of the learning assessment tools developed to capture individual abilities to meet certain minimum thresholds ('foundational skills'), which can be fielded via household surveys (**ECDI for early childhood, FLS for 7-14 years olds, PISA-D for 14-16 years olds, and mini-LAMP for adults**, and any additional instruments identified by the team in discussions with the consultant), using the insights gained from tasks (1), (2) and (3). Specifically, for each learning assessment instrument, building on the items collected in (1f), discuss the following:

   a. Identify which items identified in task (1f) appear to be amenable to phone-based administration via CATI, IVR and SMS

   b. Adaptations that need to be made to any of the items or response scales to ensure that they can be administered by phone via CATI, IVR and SMS

   c. Potential challenges in administration

This consultancy is for 25 days. The consultant will pe paid for the following deliverables:

1) Key features and evidence of validity and reliability for the suite of learning assessment instruments that can be fielded as part of household surveys (Output of tasks 1 and 2) – 5 days

2) Desk review of phone-based learning assessments (output of task 3) – 10 days

3) Feasibility study for phone-based administration of the suite of learning assessment instruments that can be fielded as part of household surveys (output of task 4) – 10 days

Qualifications

The consultant will be selected based on the following criteria:

- Master's degree (Ph.D. preferred) in topics related to educational assessment and educational policy;

- Extensive familiarity with household survey design and administration;

- Extensive familiarity with large-scale system-level learning assessments used for monitoring student learning;

- Proven experience in applying technology solutions to solve data collection challenges;

- Excellent research skills, including ability to quickly collect and synthesize information from various sources;

- Excellent writing skills and proficiency in English;

- An excellent track record of being proactive and working effectively in teams;

- Superb attention to details, and demonstrated sound organizational skills; and

- Ability to deliver assignments in a timely manner while operating under tight deadlines and juggling multiple tasks.