



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



GLOBAL
ALLIANCE
TO MONITOR
LEARNING



Measuring indicator 4.6.1 – Option paper

GAML Fifth Meeting
17-18 October 2018
Hamburg, Germany



1. Introduction

The UNESCO Institute for Statistics (UIS) has been given the mandate to monitor progress towards Target 4.6 that focuses on rates of adult literacy and numeracy skills.

Since 2005, the UIS has overseen the development and validation of the Literacy Assessment and Monitoring Survey (LAMP) to better serve the needs of developing countries in gathering informative data on literacy through household surveys.

Experience has shown, however, that there is a need for alternatives to a full LAMP assessment to reduce the operational, technical and financial burden of fielding LAMP without compromising the ability to compare results across countries and over time.

This paper identifies five alternatives/strategies which would yield significant reductions in the operational, technical and financial burden of fielding LAMP, while maintaining comparability of results. Countries should be encouraged to review their data needs and to select options best suits their needs and capacity.

Similar to implementing a full LAMP, all of these options could be pursued immediately as the tools and methodologies already exist but need minor modification.

2. Why are literacy and numeracy data needed?

The fitness for use of any data system can only be evaluated against the overall purpose for the data. As documented in **Table 1**, comparative data on the level and distribution of adult literacy and numeracy skills is needed to serve five distinct purposes, which have implications on the data collection strategy.

Comparative data on literacy and numeracy are needed for multilateral and bilateral donors to guide their policies and programmes and to monitor progress towards international and national targets, including SDG Target 4.6. It is also imperative for countries to use the data to better understanding their national situation.

Measures of literacy and numeracy need to be compared over time to determine relative needs and to track progress. The United States' National Adult Literacy Survey (NALS), International Adult Literacy Survey (IALS), Adult Literacy and Life Skills Survey (ALL), Programme for the International Assessment of Adult Competencies (PIAAC), Literacy Assessment and Monitoring Programme (LAMP) and Skills Toward Employment and Productivity (STEP) studies have established a de facto standard in which literacy and numeracy assessment results are presented on reporting scales that are divided into several theoretically-justified proficiency levels.

Table 1. Data use

Application type	General purpose	Related policy questions	Implication for data collection strategy
Knowledge generation	Identification of the causal mechanisms that link skills to outcomes	How do individuals acquire skills? How do they lose skills? How are skills linked to outcomes?	Needs longitudinal or repeated cross-sectional data with comparable measures of skills
Policy and programme planning	Planning government response to identified needs to meet social and economic goals	Which groups need skill upgrading? How many people are in need? Where is need concentrated?	Needs profile of skills for key population sub-groups
	Determination of funding levels	How much budget is needed to raise skills at the rate needed to achieve social and economic goals	Need numbers of adults with different learning needs
Monitoring	Adjustment of policies, programmes and funding levels	Are skill levels rising at the expected rate?	Need repeated cross-sectional skills measures
		Are skills-based inequalities in outcomes shrinking?	Need repeated cross-sectional skill measures for key population sub-groups
Evaluation	Formal process to determine if programmes are performing as expected and meeting their objectives	Are government programmes effective and efficient?	Need data on skills gain/loss and costs for programme participants
Administration	Making decisions about specific units: individuals, regions, programmes	What criteria are applied to determine programme eligibility?	Need results that are reliable enough to keep Type I and Type II classification errors to acceptable levels

3. Which cross-national assessments are currently available?

In an ideal world, every country would field the OECD PIAAC assessment, the World Bank's STEP assessment or the LAMP assessment as these studies provide data that meet all of the needs set out above.¹ Conventionally, these assessments include:

- Administration of an extensive background questionnaire that identifies key population sub-groups, documents the determinants of skills differences and allows to explore the impact that skill differences have on individual outcomes;
- Administration of a direct test of adult literacy and numeracy that covers the full range of skill in the population; and
- Administration of a direct test of the reading skills that support the emergence of fluid and automatic reading that characterise performance at Level 2 and below.

Given that OECD's PIAAC is designed to assess OECD countries, it might not be relevant to non-OECD countries, especially low- and middle-income countries.

UNESCO's LAMP assessment was developed to better respond to the needs of less-developed countries while maintaining the link to established proficiency scales. More specifically, the LAMP assessment:

- Includes a background questionnaire that has been adapted for use in less economically- and educationally-developed countries;
- Includes an item pool with more low-level items that provides more refinement in the lower regions of the scale; and
- Includes a filter booklet that routes less-skilled individuals to a less-demanding test and reading components.

Table 2 presents comparisons of the three assessments.

¹ Although neither PIAAC nor STEP offers results that are reliable at the individual level that is needed to support administrative or evaluative purposes, the availability of tools such as ETS' partially-adaptive literacy and numeracy assessment and Bow Valley College's fully-adaptive TOWES-Prime assessments support individual measurement on the same proficiency scales.

Table 2. Different tools and options for indicator 4.6.1

Tool option	Advantages	Disadvantage	Target countries
PIAAC	<ul style="list-style-type: none"> - Established reporting scale - Innovative implementation with computer-based mode - Produces results on both literacy and numeracy 	<ul style="list-style-type: none"> - Complex in terms of implementation - Costly - Tailored to developed and high-income countries - Computer-based implementation might be novice to developing and low-income countries - Centrally controlled and less country involvement except for the data collection 	High-income countries and those with a similar structure to OECD countries
STEP	<ul style="list-style-type: none"> - Assesses only literacy, reducing respondent burden and test time - Two-stage paper-and-pencil implementation reduces respondent burden - Results fall within the PIAAC reporting scale 	<ul style="list-style-type: none"> - Does not produce results for numeracy - Still complex in terms of implementation - Costly 	Low-income and lower-middle-income countries
LAMP	<ul style="list-style-type: none"> - Pilot countries develop items so more contextualised and in line with realities in low- and middle-income countries - Two-stage paper-and-pencil implementation reduces respondent burden - Produces results for both literacy and numeracy - Literacy domains are separated into prose and document so can further identify the type of literacy skills that the population does or does not have - Countries are involved in all stages of implementation to help build country capacity 	<ul style="list-style-type: none"> - Costly - Two-stage paper-and-pencil assessment is still complex to low- and middle-income countries, need to streamline or move to computer-based - Need to engage countries and help build capacity in all stages of implementation so it is time-consuming and takes longer to complete 	Low-income and lower-middle-income countries

In summary, all three options are costly and complex to implement. PIAAC tools may be relevant to OECD or high-income countries, but its tools are not relevant and might not be valid for low- and lower-middle-income countries. STEP tools were developed to target low- and middle-income countries, however it focuses on work-relevant skills and does not measure numeracy skills. Its premise is that numeracy skills are highly correlated to literacy skills. But it is a well-known measurement issue that using proxies can shape behavior perversely. In most countries, chemistry skills are a good proxy for physics skills, because schools that do a good job with one also do a good job with the other; but to stop testing children in physics would create perverse incentives. In low- and middle-income countries especially, it is possible to have respondents who are illiterate but have numeracy skills. LAMP worked closely with participating countries and helped build technical capacity. Especially in the era of SDG 4, countries are encouraged to take ownership of data for their own policy development. Countries are not just required to generate data for international reporting but to use the data for national monitoring.

4. Options

Data on functional literacy and numeracy are needed for low-income and lower-middle-income countries. One option is to revive the LAMP assessment programme and implement it in as many countries as possible. This would involve administering LAMP instruments to 3,000 to 5,000 respondents between the ages of 16 to 65 years per country.

The LAMP assessment, however, is paper-and-pencil-based so it does not offer any of the benefits of computer-based, adaptive assessments.

In general, PIAAC, STEP and LAMP assessments all impose high financial, technical and operational burdens on many less-developed countries. These burdens can risk the results by causing biases and errors. There is a need, therefore, to identify assessment options that serve to reduce burdens associated with fielding a skills assessment, without sacrificing the ability to compare results over time and among population sub-groups.

We have identified options to reduce operational, financial and technical burdens by using aspects of existing assessments. The dimensions to consider include:

1. Whether the number of **skills domains** is reduced. Limiting the assessment to a single skill literacy domain would **reduce the collection costs by approximately 66%**. This, however, would not collect sufficient data towards the SDGs which require the reporting of adult literacy and numeracy. The use of proxy, like self-assessment of numeracy skills, is insufficient because it could produce perverse incentives.
2. Whether to test the whole **range of skills** or limit to certain parts of the skills distribution (for example, placing individuals on the 500-point proficiency scales or classifying as being above or below an educationally- and economically-material threshold e.g. above or below the Level 2 or 3 threshold.)
3. Whether the assessment will be conducted as a dedicated study or added on to an existing study. The latter option reduces the financial burden of fielding an assessment and allows for more efficient sample allocations.

4. Whether the assessment will be designed to provide direct point estimates of skills distributions or to support the generation of indirect synthetic estimates. Since the latter approach requires a smaller sample size, it would reduce collection costs by 30% to 50%.
5. Whether the assessment will be paper-and-pencil or computer-based. Computer-based options yield significant reductions in the operational, technical and financial burdens of fielding an assessment. Collection costs are **reduced by 40%** while yielding **individually reliable results across the entire skills distribution**. Some child-level assessments, such as EGRA, have moved entirely to tablet-based (though with a paper stimulus), for cost, accuracy and speed.
6. Whether to keep an international coordinating agency or to make measurement tools and associated implementation packages available for every country to decide how to administer, keeping a quality assurance mechanism in place.

These options are summarised in **Figure 1**.

Figure 1. Options for SDG Indicator 4.6.1

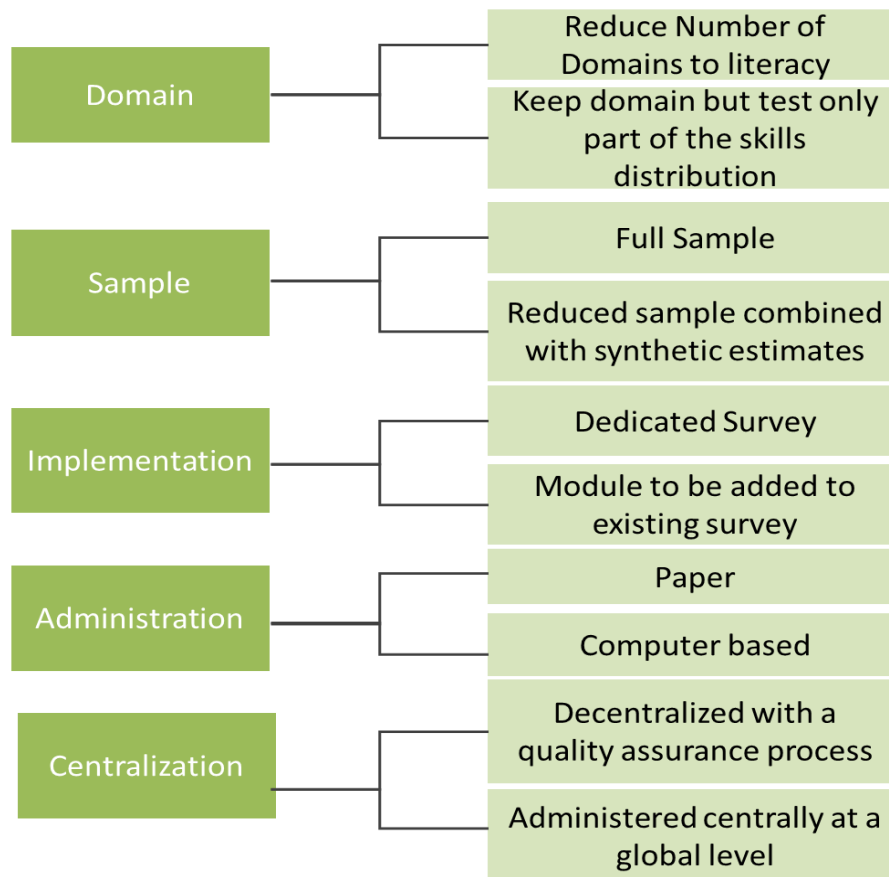


Table 3 presents an analysis of the advantages and disadvantages of various options to reduce operational, financial and technical burdens by using aspects of existing assessments. A more detailed analysis is presented in the **Annex**.

Table 3. Advantages and disadvantages of different options to measure SDG Indicator 4.6.1

Option	Advantages	Disadvantages	Intended coverage
Domain			
Reduce the number of skills domains	<ul style="list-style-type: none"> - Reduces the cost and response burden associated with obtaining data on skills because prose literacy and numeracy are very highly correlated (95%) 	<ul style="list-style-type: none"> - No results for population who might be illiterate but possess numeracy skills - Could produce undesirable, unintended consequences, such as minimising numeracy in instructional programmes 	<ul style="list-style-type: none"> - Produce literacy results only
Reduce the scope of the test to particular skills and classify adults above a threshold	<ul style="list-style-type: none"> - Reduces the amount of testing time and, thereby, the cost and operational burden by perhaps an additional 25% 	<ul style="list-style-type: none"> - A high share of developing countries would see the majority of their population classified below the threshold - The LAMP item pool does not contain sufficient items around the specified threshold so additional items would have to be developed and validated. Therefore other existing item pools, like the TOWES-Prime test, could be a potential to explore - The threshold would have to be picked in advance and would presumably have to be global in order to effect the cost-savings, which would require agreement in-advance among the relevant countries 	<ul style="list-style-type: none"> - Allow sampled respondents to be classified into performance levels based on outcomes of assessment
Sample options			
Reduce sample size to a purposive sample and apply 'big data' analytics	<ul style="list-style-type: none"> - Requires significantly smaller sample sizes, so reduces the burden of obtaining skills profiles - Uses multivariate statistical inference to produce point estimates based on correlates of literacy 	<ul style="list-style-type: none"> - Slightly broader confidence intervals on the final estimates 	<ul style="list-style-type: none"> - Selective sample of respondents are assigned cognitive modules and use relationship to generate covariate to project
Modality of implementation			

Administer to sub-sample of an existing household survey	<ul style="list-style-type: none"> - Reduces the response burden significantly because the base sample would already collect most of the background variables needed to describe skills levels - Reduces the cost of fielding a skills assessment since the fixed costs of fielding an assessment would already have been covered 	<ul style="list-style-type: none"> - Increases response burden with the added module which has a negative impact on response rates for some population sub-groups - Non-response may increase, introducing bias into the skills estimates - This option also involves a level of coordination of activities and cooperation between agencies that may be difficult to achieve 	<ul style="list-style-type: none"> - A representative sub-sample of respondents of existing household-based surveys. This saves on sampling
Modality of administration			
Administering a fully-adaptive computer-based (web-based) skills assessment	<ul style="list-style-type: none"> - Reduces average test duration by at least 40%, so reduces collection costs and response burden by roughly the same magnitude - Yields proficiency scores that are reliable at the individual level - Tests are scored in real time - Eliminates the need for post-assessment data capture and processing so that analysis can begin immediately 	<ul style="list-style-type: none"> - Test suites require stable access to the Internet, if web-based, and a laptop or tablet. Although it may be accessed through a local hotspot linked to a cell phone network, if available, the data can be downloaded at the end of the day to a laptop or tablet when the assessors reach a spot with 3G capabilities 	<ul style="list-style-type: none"> - Adaptive nature of the test allows coverage of both literacy and numeracy in a shorter timeframe yet produces more reliable results
Central coordination			
Administered centrally by an organization	<ul style="list-style-type: none"> - Centralised agency ensures quality assurance in data collection and processes for analysis - Enables production of quality data - If properly coordinated, countries will be able to work with the central agency and learn from them 	<ul style="list-style-type: none"> - Development will be conducted centrally - Decisions made with limited inputs from countries for some processes 	<ul style="list-style-type: none"> - Quality assurance ensures quality data for reporting

Measurement tools plus implementation package developed to ensure comparability	Measurement tools and implementation package developed to adhere to SDG 4 reporting requirements Tools and package available in major languages (English, French and Spanish) Guidance on the use of tools and packages	Tools and packages might not incorporate country contexts thus may need adaptation	Tools and packages available for adaptation
--	---	--	---

5. What is needed for implementation?

Any of these options require a standardised process to implement. Key inputs in this process include:

1. Module on cognitive items. Cognitive items currently exist through LAMP and PIAAC. While PIAAC modules target developed economies, LAMP modules target low- and middle-income countries. LAMP items have been field-tested in eleven countries² and are operationalised in five countries³. LAMP cognitive items have been translated into Arabic, Dari, French, Lao, Mongolian, Pashto, Spanish, Vietnamese, and five African languages: Fulfulde, Hausa, Kanuri, Tamasheq and Zarma.
2. Preparation of a guide that provides specifications on how to administer the assessment. The activities would depend on the collection mode: paper-and-pencil or computer-based. The latter process is far less operationally- and technically-demanding.
3. Prepare translation and adaptation guide for the test items and reading components. Procedures for the former are well established. Research shows that reading components vary significantly from language to language, so these need to be carefully developed by linguists for each new language.
4. A sampling guide needs to be developed that reflects whether the assessment is to be conducted as a stand-alone survey, as an add-on to an existing survey or as purposive addition to an existing sample.
5. A data capture and processing guide need to be prepared. For computer-based assessments, this will be limited to the processes needed to code and capture open-ended and other specific responses and the weights of each respondent in the database.
6. Data analytics.

² Afghanistan, El Salvador, Jordan, Kenya, Mongolia, Morocco, Niger, Palestine, Paraguay, People Republic of Lao and Vietnam

³ Jordan, Mongolia, Palestine, Paraguay and People Republic of Lao

Quality assurance

Experience with IALS, ALL, PIAAC, STEP and LAMP suggests a need for rigorous quality assurance and support throughout implementation, otherwise countries face a significant risk of introducing bias into the estimates.

There are six elements to a quality assurance regime:

- i. Clear specifications for all activities.
- ii. Explicit standards to be met.
- iii. Preparation of a national planning report that details national implementation plans.
- iv. Training on key activities, including translation/adaptation, collection, data processing and analysis.
- v. Analysis of compliance evidence by the international team.
- vi. Third party observation of collection.

The LAMP assessment has a full set of documentation for a paper-and-pencil implementation.

Legal framework

Experience with IALS, ALL, PIAAC, STEP and LAMP suggests a need for two sets of undertakings:

- i. An MOU with implementing agencies, such as the World Bank, UNICEF, regional development banks and bilateral donors, that set out roles, responsibilities and expectations.
- ii. MOU with participating countries that specifies roles, responsibilities and expectations, including access to microdata and the publication of results.

Drafts of both documents were developed for LAMP and these could be adapted for current use.

6. Timeline

Each step involves time. Depending if the assessment is pen-and-pencil or computer-based, a full cycle could last from 11 to 24 months (see **Table 4**).

Table 4. Timeline for various activities

Activity	Timeframe required
Develop or modify existing materials to international level	
Paper-and-pencil cognitive modules	One to two months based on available tools
Guidelines for implementation	One to two months based on available material
Adapt re-design tools to computer environment	Two to three months
Translation and adaptation of updated materials	Two to three months
Data processing software for re-design modules	One to two months
Quality assurance	One month
Legal framework development	Three months
Implementation at country level	
National reporting	Three months
Data collection	Four months for paper-and-pencil assessment
	One to two months for computed-based assessment
Data processing	Four months for paper-and-pencil assessment
	One month for computed-based assessment
'Big data' (synthetic) estimates	One month

7. Cost of administration

Experience suggests that data collection and processing costs vary significantly by country in ways that are difficult to predict. Thus, the only way to get accurate estimates for any given country is to have them produce a national planning report and associated cost estimates. **Table 5** provides indicative cost estimates for each of the proposed options for seven rates of interviewer pay.

Computer-based options are systematically less expensive than paper-and-pencil options because they involve less interviewer training, reduce the average time taken to complete an interview and eliminate data processing and scoring costs. They also reduce the amount of time that national teams need to function by as much as five months.

On another level, computer-based options are more efficient because they shift expenditures from the act of collection to the analysis of results.

In addition to the country-level cost as presented in Table 3, an initial investment for development, which includes updating or modifying existing material, will amount to approximately US\$230,000.

Table 5. Indicative costs for various collection options at varying interviewer wage rates

		Interviewer wage rate						
		\$5	\$6	\$7	\$8	\$9	\$10	\$15
Total costs (US\$)	1,500 cases, web-based	157,550	166,370	175,190	200,110	208,930	233,850	227,950
	3,000 cases, web-based	236,650	254,290	271,930	305,670	323,310	357,050	445,250
	5,000 cases, web-based	320,550	348,970	377,390	421,910	450,330	494,850	636,950
	1,500 cases, web-based, 50 countries	7,877,500	8,318,500	8,759,500	10,005,500	10,446,500	11,692,500	13,897,500
	3,000 cases, web-based, 50 countries	11,832,500	12,714,500	13,596,500	15,283,500	16,165,500	17,852,500	22,262,500
	5,000 cases, web-based, 50 countries	16,027,500	17,448,500	18,869,500	21,095,500	22,516,500	24,742,500	31,847,500
	3,000 cases, paper-and-pencil	420,880	469,516	518,152	598,988	647,624	728,460	971,640
	1,500 cases, paper-and-pencil	252,528	281,710	310,891	359,393	388,574	437,076	582,984
	3,000 cases, paper-and-pencil, 50 countries	21,044,000	23,475,800	25,907,600	29,949,400	32,381,200	36,423,000	48,582,000
	1,500 cases, paper-and-pencil, 50 countries	12,626,400	14,085,480	15,544,560	17,969,640	19,428,720	21,853,800	29,149,200

8. Recommendations

The mini-LAMP will provide an alternative to meeting the needs of low-income countries when assessing literacy. Based on the cost options, technical feasibility and national needs, the UIS proposes the following recommendations:

- Adapt the existing LAMP item pool, background and cognitive tools, administration and implementation guides, and quality assurance package that are targeted to low-income countries.
- Create short modules for both literacy and numeracy domains.
- Use a computer-based platform that has been fully validated and tested to host short modules.
- Explore existing items that could be mapped to the content framework to expand the item pool.
- Take advantage of the existing implementation platform within household surveys and attach a short literacy module to it.
- Negotiate with implementing agencies on collaboration and support for assessment field administration, observed interviews and data processing.

In summary, the UIS considers that a paper-and-pencil assessment could be launched rapidly, as administration and implementation guidelines (i.e. interviewer training materials, procedures and manuals); tools (i.e. background questionnaires and assessments); the quality assurance package; and legal framework template (i.e. National Planning Report) are already available.

The UIS will need to develop the computer platform, negotiate and work with implementing partners to ensure the implementation platform is ready and with technical partners to ensure that quality assurance procedure is in place.

Annex. Advantages and disadvantages of different options to measure SDG Indicator 4.6.1

1. By variants in **domain** administration
 - a) Reducing the number of skills domains assessed;
 - b) Reducing the scope of testing within each domain to a reduced part of the skills in the framework. By administering a skills assessment with the sole purpose of classifying adults above or below a key threshold, e.g. above or below 275 points, the threshold between literacy Levels 2 and 3.
2. By **sample** administration changes
 - a) Administering a skills assessment to a purposive sample of respondents that provides estimates of respondents with a specific characteristic, e.g. women in a rural location, and then using the relationship to assign the probability of a proficiency level to a given population;
 - b) By administering a skills assessment to a sub-sample of respondents in an existing household survey.
3. Computer-based or paper-and-pencil assessments

1. Domain

a) Reducing the number of skills domains assessed

In this option, the burden of administering direct assessments of prose literacy, document literacy and numeracy and reading components is reduced by limiting the number of skills domains assessed.

Advantages

Reducing the number of skills domains lowers significantly the costs and response burden associated with obtaining these data, without compromising the usefulness of the data for formulating policy and programmes and for monitoring. This approach works because prose literacy and numeracy are highly correlated (95%), so a measure of one generally gives the same results for the other. Similarly, numeracy skills are highly dependent upon literacy skills. This strategy is used by the World Bank in the STEP programme.

Disadvantages

No numeracy results will be available if respondents are not tested in numeracy. The available adaptive test module(s) will provide users with the choice of single or multiple domains.

b) Reducing in the scope of the testing to a range of skills

Operational, financial and technical burdens are reduced by administering a skills assessment with the sole purpose of classifying adults above or below a key threshold, e.g. above or below 275 points, the threshold between literacy Levels 2 and 3.

This option involves concentrating test items around a score threshold that is educationally and economically meaningful.

In reality, the appropriate proficiency cut-off would reflect the level needed for a country to meet their economic and social goals. Thus, the appropriate threshold would vary given a country's level of educational, social and economic development.

In practice, the cut-off between literacy Levels 2 and 3 in IALS, PIAAC and STEP could be adopted as a universal standard, as Level 3 is the level at which processing moves from routine recall to fluid problem-solving in the pre-frontal cortex. It is the latter skill level that allows adults to solve unfamiliar problems efficiently through applying their fluid and automatic reading skills.

The UIS would need to select a testing partner with sufficient items and supervise the development and validation of the threshold test in each new language.

Advantages

This option has the advantage of further reducing the amount of testing time and, thereby, the financial and operational burdens by perhaps an additional 25%.

Disadvantages

A high share of developing countries would see the majority of their population classified below the threshold. Over time, attention will need to be focused on the rate at which this share is decreasing. Given the fact that most of the countries would start well below the 275-point threshold, one would expect to observe very little change over the short term.

The LAMP pool does not contain sufficient items around the specified threshold, so additional items would have to be developed and validated for a paper-and-pencil implementation. The TOWES-Prime test includes sufficient items and is adaptable enough to support this type of test.

2. Sample

a) Purposive or convenience sample of a survey that provides estimates of the probability of being proficient

In this option, a purposive sample is selected capturing as much variation in the population sub-groups as possible. The sample would then represent the probability of different proficiency levels. In most cases, this would involve estimating the probabilities for different education levels within a matrix of age group, gender, employment status and, in some countries, having a mother tongue other than the language of instruction. These probabilities would then be used to perform a multiple imputation of proficiency scores onto a target survey that contains the same demographic co-variates. This option could involve administration of a paper-and-pencil assessment or a web-based, fully-adaptive

assessment. The use of a web-based adaptive assessment would reduce the financial and operational burdens by at least 40%.

The UIS would provide advice on the design of the purposive sample and would run the multiple imputation that is used to generate synthetic estimates.

Advantages

This option requires significantly smaller sample sizes than needed to generate point estimates of the numbers of adults at each proficiency level, so it reduces the operational and financial burdens of obtaining skills profiles without compromising the ability to compare results. Sample sizes of 1,500 cases are generally sufficient rather than 3,000 to 5,000 cases needed for a conventional assessment. One could expect a 30% to 50% reduction in variable collection costs as a result of choosing this option.

Disadvantages

None.

b) Implementation: Skills survey to a sub-sample of respondents of an existing household survey

In this option, a sub-sample is selected from respondents to an existing survey, such as MICS, DHS or a national Labour Force Survey.

This option could involve administering a paper-and-pencil assessment or a web-based, fully-adaptive assessment. The use of a web-based adaptive assessment would reduce the financial and operational burdens by at least 40%.

In this case, the UIS would need to partner with the sponsoring multilateral, bilateral and/or national agencies associated with the existing survey, e.g. UNICEF for MICS, the World Health Organization (WHO) for DHS, the World Bank for STEP or the national statistical offices for national surveys.

The UIS would then adapt the instruments and associated LAMP documentation, vet the National Planning Reports, deliver training, provide national teams with basic analysis of skills distributions, determinants and the link to outcomes, and take the lead on publishing comparative results to report on SDG Indicator 4.6.1.

National implementing agencies would assume responsibility for all aspects of implementation and the publication of national results.

Advantages

This option would reduce the response burden significantly because the base sample would already collect most, but not all, of the background variables needed to describe skills levels. Key variables on skills use would need to be collected.

It would also reduce the cost of fielding a skills assessment since the fixed costs of fielding an assessment – building a sample frame, recruiting and training interviewers, and building a data capture and processing system – would already have been covered.

Disadvantages

The sole disadvantage of this option is that it may increase response burden enough to have a negative impact on response rates for some population sub-groups. Non-response may increase enough to introduce bias into the skills estimates.

This option also involves a certain level of coordination of activities and cooperation between agencies that may be difficult to achieve

3. Administration: Migrate to a fully-adaptive, web-based skills assessment

Existing assessment programmes rely on tests that are not adaptive or only mildly adaptive. This option would involve replacing partially-adaptive tests with fully-adaptive, web-based tests. The LAMP test could be mounted on a test delivery platform, but the LAMP item pool does not contain sufficient items to support a fully-adaptive test. The Education Testing Service (ETS) has developed a partially-adaptive test in partnership with the OECD and the European Union (EU); the Government of Canada funded the development and validation of a test delivery platform and a suite of fully-adaptive tests of prose literacy, document literacy, health literacy, numeracy and reading components called TOWES -Prime. Both test suites are already available in multiple languages.

The UIS would need to select a testing platform and associated item pool, provide guidelines for translation and adaptation, and develop quality assurance procedures.

Advantages

Using a fully-adaptive assessment reduces average test duration by at least 40%, so it reduces collection costs and response burden by about the same magnitude.

Fully-adaptive tests also yield proficiency scores that are reliable at the individual level and that are equally reliable no matter where an individual is on the proficiency distribution. In contrast, tests that are not fully-adaptive yield sufficiently-reliable results around the average difficulty test items and far less reliable results in the tails of the proficiency distribution.

The use of a fully-adaptive, web-based test also eliminates the need for any post-assessment data capture and processing, so that analysis can begin immediately. This means more of the available budget can be devoted to extracting policy value from the data.

Disadvantages

Both of the available test suites require stable access to the Internet and a laptop. Under normal circumstances, sufficient bandwidth can be accessed through a local hotspot linked to the cell phone network.