unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

WG/GAML/11/5.1

PAL Network's ICAN-ICARe: Alignment with Global Reporting Criteria for SDG 4.1.1(a)

Version 4 - 9 December 2024

Global Alliance to Monitor Learning

Introduction

The People's Action for Learning (PAL) Network is a South-South partnership of organizations working to improve foundational literacy and numeracy (FLN) for children through citizen-led, large-scale assessments and community-based education initiatives. PAL Network operates across 15 countries in Africa, Asia, and Latin America, empowering local actors to generate data that informs education policies and holds governments accountable for delivering quality education.

One of PAL Network's flagship initiatives is the **Large-Scale Assessment (LSA) project**, which aims to produce globally comparable, yet contextually relevant, data on children's foundational learning. This project has led to the development of two key assessment tools: the **International Common Assessment of Numeracy (ICAN)** and the **International Common Assessment of Reading (ICARe)**. These tools are designed to assess children's proficiency in numeracy and reading, respectively, using a cost-effective, paper-based approach that is suitable for diverse settings, including remote and low-resource environments. The ICAN-ICARe tools are aligned with the Global Proficiency Framework (GPF) to enable reporting on Sustainable Development Goal (SDG) indicator 4.1.1(a), which tracks the proportion of children in grades 2 and 3 achieving at least a minimum proficiency level in reading and mathematics.

The **Global Criteria for Reporting on SDG 4.1.1(a)**, developed by the Global Alliance to Monitor Learning (GAML), outlines the technical requirements that assessments must meet to ensure global comparability, reliability, and construct validity. This includes alignment with the GPF, sufficient sampling precision, and inclusive assessment design. These criteria are critical for ensuring that assessments can provide accurate, equitable data on children's foundational learning outcomes. For PAL Network, meeting these criteria is essential to ensuring that the ICAN-ICARe assessments contribute meaningfully to the global education discourse and advocacy efforts aimed at improving foundational literacy and numeracy.

The following sections provide an overview of where PAL Network currently stands in relation to the global reporting criteria. A detailed analysis highlights the strengths of the ICAN-ICARe tools and the areas where additional progress is needed to fully meet the requirements by December 2025.

Status of PAL Network's ICAN-ICARe Alignment with Global Reporting Requirements (as of January 2025), with Progress Projections for December 2025

Table 1 below provides a summary of PAL Network's ICAN-ICARe assessment progress in meeting the global reporting requirements for SDG 4.1.1(a). The criteria are evaluated across three categories: **currently met**, **expected to be met by December 2025**, and **areas of concern requiring additional focus**.

Table 1: Status of PAL Network's ICAN-ICARe Alignment with Global Reporting Requirements

Criterion	Currently met	Expected to be met by December 2025	Areas of concern requiring additional focus	Not Applicable
1	1.1a (R) 1.2a (R) 1.3a (R) 1.4a (R) 1.7a (R) 1.1a (M) 1.2a (M) 1.3a (M) 1.4a (M) 1.8a (M)	-	0	0
2	2.1 2.3 2.4 2.5 2.7	2.8 2.10 2.12	2.6	0
3	3.1 (S) 3.4 (S) 3.8 (S)	3.2 (S) 3.3 (S) 3.5 (S) 3.9 (S) 3.10 (S) 3.11 (S)	0	3.6 (S) 3.7 (S)
4	4.1 4.3 4.4 4.5 4.9 4.8 4.10	4.11	0	0

Criterion	Currently met	Expected to be met by December 2025	Areas of concern requiring additional focus	Not Applicable
	4.12			
5	5.9	5.1 5.2 5.3 5.4 5.5 5.6 5.7	0	0
6	0	6.1 (PCM) 6.2 (PCM) 6.3 (PCM) 6.4 (PCM) 6.5 (PCM) 6.6 (PCM) 6.7 (PCM) 6.8 (PCM) 6.9 (PCM)	0	6.1 (SL) 6.2 (SL) 6.3 (SL)
7	0	7.3 7.4	7.1 7.2	7.5
Total	27	28	3	6

Out of the seven criteria, Criterion 1 is fully met with no anticipated concerns or gaps. Criterion 2 is partially met, with progress needed in one area to fully meet expectations. For Criteria 3 and 4, significant progress is expected, with six and one additional elements, respectively, expected to meet compliance by the end of 2025. Criterion 5 shows substantial room for improvement, with only one element currently meeting requirements and seven elements targeted for completion.

Criteria 6 and 7 reflect areas requiring the most development. Criterion 6 has no current compliance but is expected to meet nine elements by December 2025. Criterion 7 is notably flagged, with two areas of concern that may pose challenges to achieving full alignment despite planned improvements. This table underscores both current strengths and the roadmap for future alignment with global standards, highlighting specific areas where additional resources or interventions may be needed.

The following section describes the status and future progress for each criterion.

Criterion 1: Alignment to the MPL and content validity

Current Status:

PAL Network meets the requirements for this criterion across both reading and numeracy assessments. The Assessment Blueprint for the International Common Assessments of Numeracy (ICAN) and Reading (ICARe) demonstrates a strong alignment with the Global Proficiency Framework (GPF) and construct validity requirements, meeting and, in several areas, exceeding the minimum criteria. The blueprint outlines 30 items for reading and 36 items for numeracy, surpassing the minimum required 20 score-points per assessment. The design reflects comprehensive coverage across key domains and subconstructs for both reading and numeracy, ensuring that the assessments accurately measure foundational learning as stipulated by SDG 4.1.1(a).

For ICARe, all three critical domains—Listening Comprehension, Decoding, and Reading Comprehension—are represented, with balanced item allocations that address both retrieval and inferencing tasks. The decoding section meets the requirement of at least 10 score-points, while the listening and reading comprehension sections include sufficient score-points to ensure validity in measuring students' proficiency levels. In numeracy, ICAN includes five key domains, with a notable emphasis on Number and Operations, supplemented by items in Measurement, Geometry, Statistics, and Algebra. The breadth of subconstructs in both reading and numeracy assessments reflects thoughtful alignment, covering a wide range of skills relevant to grades 2 and 3.

The blueprint's design includes both open-ended and closed-ended items, ensuring a variety of response formats that cater to different cognitive demands. Detailed scoring rubrics and coding instructions are provided for standardized administration, supporting consistent interpretation of children's responses. Additionally, the blueprint emphasizes rigorous training for enumerators to maintain administration fidelity, reinforcing the reliability and replicability of the data. Sampling procedures are designed to capture representative data across 15 countries, demonstrating a commitment to producing globally comparable results.

The assessment blueprint also addresses contextual considerations, acknowledging the impact of assessment language versus home and instruction languages on children performance. Plans to pilot inclusive modules—such as adaptations for children with special needs and provisions for social and emotional learning (SEL) and education in emergencies (EiE)—reflect a forward-looking approach that promotes accessibility and equity. These components align with the inclusivity goals outlined in the GPF and enhance the assessment's applicability in diverse educational contexts.

Table 2: General Criteria for Alignment to the GPF and Construct Validity

Criteria	Minimum Requirement	LSA Assessment Blueprint	Comparison
Number of Items	Minimum 20 score-points for reading and numeracy	30 items in ICARe, 36 in ICAN	Exceeds the minimum requirement for both numeracy & reading.
Domains for Reading (ICARe)	Listening Comprehension, Decoding, Reading Comprehension	Includes all three domains with a breakdown of items per domain	Fully aligned; all domains are represented as required.
Domains for Numeracy (ICAN)	Number Operations, Measurement, Geometry, etc.	Includes five domains, with significant emphasis on operations	Fully aligned; includes domains beyond minimum requirements.
Construct Breadth (Reading)	At least 4 subconstructs covered in Decoding and Comprehension	Covers 5 subconstructs in Grade 2, 6 in Grade 3	Exceeds the minimum number of subconstructs for both grades.
Construct Breadth (Numeracy)	At least 4 subconstructs covered across foundational topics	Covers 11 subconstructs across five domains	Strongly aligned; more subconstructs than the minimum.
Construct Validity	Items must correspond to skills relevant for Grades 2/3	Explicitly mapped to GPF constructs and minimum proficiency levels (MPLs)	Fully aligned; item descriptions correspond to GPF descriptors.

The assessment blueprint is attached as Annex A.

Future Progress:
No further actions are required for this criterion as all necessary requirements are already met.

Areas of Concern:
There are no areas of concern for Criterion 1.

Criterion 2: Item content and quality

Current Status:

The ICAN-ICARe assessments have been developed through a robust process involving subject matter experts (SMEs) and local stakeholders. Items have been reviewed for contextual and cultural appropriateness, and scoring guides align with the constructs they are designed to measure. The development team follows established psychometric principles, and where possible, items undergo iterative revisions before live assessment.

We have a planned field test for Q1-2025 which speaks to the Criterion 2.6. We have designed a detailed sampling approach to ensure a robust evaluation of the assessment tools. The purpose of the field test is threefold: to evaluate the clarity, relevance, and cultural appropriateness of the questions across different language groups; to identify biases or difficulties that could impact fairness; and to validate the psychometric properties of the tools, ensuring their reliability and suitability for broader application.

In terms of sample size, we have considered statistical power and precision to detect meaningful differences within and between language groups. Our technical partner, ACER, recommends a minimum of 30 responses per group per condition for reliable results. For ICAN-ICARe, a "group" refers to each language requiring testing, and "conditions" include variables such as rural vs. urban settings and gender differences. Accordingly, we plan to engage at least 100 households with children aged 5–16 years for each language group under both rural and urban conditions. This will involve sampling from 10 villages or enumeration areas (EAs) per language group—5 rural and 5 urban—covering 20 households per village or EA, totaling 200 households per language group.

This structured approach ensures sufficient variability within the sample to capture potential biases and contextual challenges, providing actionable insights for refining the tools before nationwide implementation. By adopting this strategy, we aim to comprehensively assess the assessment tools' efficacy and fairness while maintaining a balance between logistical feasibility and robust data collection.

Future Progress:

By December 2025, PAL Network will strengthen its psychometric analyses, incorporating item response theory (IRT) to ensure accurate data on item performance. Efforts will focus on improving post-field trial reviews to finalize items based on difficulty, discrimination, and inter-rater reliability measures.

Areas of Concern:

One of the key challenges lies in Criterion 2.6, which requires field testing of items on a representative sample of the learner population before live administration. Household-based assessments, such as ICAN-ICARe, aim to collect nationally representative data during the live assessment itself. Conducting a full-scale pilot that mirrors the national sampling approach presents significant logistical and financial constraints, as household assessments involve travel across diverse and remote regions.

To address this concern, we propose the following considerations to align with the intent of the field-testing requirement without requiring nationally representative pilot samples:

1. Targeted Regional Pilots: Field tests could be conducted in a selected number of socio-demographically diverse regions to capture potential issues related to language, phrasing, or context. This approach provides valuable data on item performance without the need for national coverage.

2. Cross-Country Item Validation Workshops: Convening expert panels from participating countries to review item translations and cultural appropriateness before the live assessment can help identify potential challenges early in the process.

3. Iterative Pre-Testing: Small-scale pre-tests in a few key locations, followed by rapid item revisions, can improve the quality and clarity of items before national implementation. This strategy reduces the risk of poor item performance without requiring a nationally representative pilot.

These approaches balance the need for psychometric rigor with the realities of household-based assessments. By focusing on smaller but diverse pilots and expert validation, PAL Network can continue to ensure the quality of its assessment items while adhering to resource constraints.

Criterion 3: Population coverage and sampling

Current Status:

The target population for reporting is clearly defined, and the sampling frame is documented. The sampling design ensures stratification by key variables such as rural/urban and language.

Future Progress:

Documentation and application of sampling weights will be completed by 2025 to ensure accurate national estimates. While the PAL assessments overall response rate already meets the 70% requirements, and we do publish reports with gender-disaggregated data, further efforts will focus on enhancing documentation to maintain and demonstrate compliance across all reporting criteria.

Areas of Concern:

There are no areas of concern for Criterion 3.

Criterion 4: Assessment administration and data custodianship

Current Status:

PAL Network has a standardized administration manual and processes for selecting, training, and qualifying enumerators, district coordinators (DCs) and project management teams (PMTs). Quality assurance plans include monitoring during the survey, desk and field rechecks, and recording incidents of improper administration.

Future Progress:

By December 2025, further checks for inappropriate administration will be strengthened, and a more robust system for informed consent will be implemented.

Areas of Concern:

No significant concerns, though detailed reporting on corrective actions for misadministration is an area for continuous improvement.

Criterion 5: Reliability

Current Status:

Some reliability analyses, such as Cronbach's alpha, are performed. Plans to review differential item functioning (DIF) and ensure a kappa coefficient above 0.8 for inter-rater reliability have been outlined.

Future Progress:

Full implementation of IRT-based reliability testing, achieving a reliability coefficient of at least 0.8, and thorough post-administration analyses will be completed by 2025.

Areas of Concern:

None.

Criterion 6: Benchmark-based linking to the MPL

Current Status:

Criterion 6 outlines the requirements for linking of assessments to the Minimum Proficiency Levels (MPLs) using either statistical linking (SL) method or pairwise comparison method (PCM), as well as guidelines for conducting psychometric analysis and ensuring appropriate sample representation. At present, PAL Network has not yet fully implemented this work for its assessments; however, this is an area of planned work.

- PCM (Pairwise Comparison Method) Requirements (6.1 - 6.9):
  PAL Network's assessments are expected to meet these criteria by the end of 2025. This includes ensuring that:

1. Participants are representative and meet the participation criteria (6.1, 6.2).
2. Responses that do not fit the model are removed with clear justifications (6.3, 6.4).
3. Reliability indices meet thresholds (e.g., pairwise scale reliability ≥ 0.75 for 6.5).
4. Differential item functioning (DIF) analysis is performed, and problematic items are removed (6.6).
5. The dis-attenuated correlation for linked items meets required standards (6.7).
6. Mean scores for each evaluation section meet minimum scores (6.8).
7. An impact analysis workshop is held to validate the linking process (6.9).

Future Progress:

By December 2025, PAL Network aims to implement the pairwise comparison method (PCM) for linking its ICAN-ICARe tools to MPLs. This will involve conducting workshops, finalizing protocols for linking, and ensuring that psychometric analyses meet reliability and representativeness standards.

Areas of Concern:

No significant concerns. However, the requirement for consistent linking data across multiple cycles might present a challenge, particularly due to the technical demands of maintaining stable item performance over time.

Criterion 7: Maintaining standards over time

Current Status:

Progress is ongoing in strengthening protocols for the reuse of public and restricted items to ensure consistency and transparency. While some items have been reused across different assessments, efforts are focused on formalizing comprehensive public release guidelines and procedures that align with global best practices.

Future Progress:

By 2025, a process for managing the public release and reuse of items, as well as equating assessments over time, will be formalized.

Areas of Concern:

One of the key concerns for PAL Network regarding Criterion 7 relates to the public domain status of assessment items. Criterion 7.1 allows for the reuse of items that are not in the

public domain in multiple live test administrations, while Criterion 7.2 mandates that any items released to the public domain cannot be reused in future assessments.

Historically, PAL Network has prioritized transparency and accessibility by publishing all assessment tools, including the items, administration processes, and the data collected, under open-access principles. This practice aligns with PAL's commitment to fostering accountability and knowledge sharing across its global network. However, the global reporting requirement restricting the reuse of publicly available items presents a significant challenge for PAL's household-based assessments, which are distinct from traditional school-based assessments.

Unlike school-based assessments that often prioritize secure item banks to prevent test coaching or item exposure, household-based assessments focus on community engagement and public dissemination of data to promote advocacy. The reuse of some items in successive rounds, particularly anchor items for longitudinal analysis, is an essential practice for ensuring comparability across cycles. Restricting this reuse could hinder PAL's ability to maintain consistency while upholding its open-access approach.

Given the distinct objectives and design of household-based assessments, a key consideration for policymakers could be whether this restriction is necessary for non-school-based assessments. Household-based assessments could be granted flexibility to publish their tools while still allowing partial reuse of items, especially where there is no risk of "teaching to the test." This distinction would enable PAL Network to preserve its best practices of open access and transparency without compromising the technical rigor required for global reporting.

In summary, 7.1 and 7.2 remain areas of concern due to the potential conflict with PAL's values and approach to open-access publication. A tailored approach for household-based assessments that accounts for their unique role in community-level advocacy and data generation may be a feasible solution for maintaining both transparency and compliance.

PAL Network's ICAN-ICARe:
Alignment with Global Reporting Criteria for SDG 4.1.1(a)

Email:

uis.information@unesco.org

uis.director@unesco.org

uis.unesco.org

@UNESCOstat