Koloso Working Paper WP/2026/01
Evidence and impact series

Assessment at scale: What a national competition tells us about Koloso's assessment methodology

Can low-cost, continuous curriculum-aligned digital assessment complement or replace periodic large-scale testing? Preliminary evidence from 107 schools in Zambia.

James Blewett, Co-Founder and CEO, Koloso Ltd | March 2026 | koloso.app/evidence

Abstract

On 6 March 2026, Koloso hosted the Mini CEO Round 2 competition on its platform, involving 107 schools across Zambia. The competition used a combination of quiz questions created specifically for the competition (about the Mini CEO book narrative contents) and drawn from Koloso's own curriculum-aligned question bank, covering social and commercial arithmetic from the primary mathematics syllabus. The distribution of school scores across the three-game competition produced a near-normal bell curve. This working paper examines what that result tells us (and what it does not) about the validity of Koloso's assessment methodology. It argues that the result is a meaningful proof of concept, identifies the limitations of the current evidence, and proposes a low-cost, high-value next step: offering free Koloso access to participating government schools to run alongside Zambia's forthcoming UNICEF/FCDO national learning assessment programme, creating the conditions for a correlation study that would substantially strengthen the evidence base for curriculum-aligned digital assessment as a complement, and eventually a partial substitute, for periodic large-scale testing.

1. Background and context

1.1 The Mini CEO programme

Mini CEO is an independent educational initiative that produces a series of books for upper primary school children, set in African contexts and designed to develop entrepreneurial thinking and financial literacy. Koloso has a partnership with Mini CEO to host the programme's national competition on its platform and to develop a bank of age-appropriate, largely curriculum-aligned quiz questions drawing on the narrative contents of the books and the social and commercial arithmetic strand of the Zambian mathematics syllabus.

Mini CEO Round 2 took place on 6 March 2026. Koloso hosted the competition on its platform, with 107 schools participating across Zambia - a mixture of government and private schools, with a skew towards the private sector in this round. Each school entered as a team and competed across three games of 10 questions with a time limit of 2 minutes per game, with their scores analysed and presented in real time back through an adapted version of the Koloso teacher's dashboard and the Koloso leaderboard (accessible to all participating schools). The school teams were supervised by their own teachers - with some submitting videos of their play.

1.2 How Koloso's assessment normally works

It is important to note that the Mini CEO competition format (team-based, three games, competitive) differs from Koloso's standard formative assessment approach. In Koloso's normal deployment, every student has their own unique login and faces an individual, personalised sequence of questions that are grade-specific and curriculum-aligned to the national syllabus of their country (currently Zambia, South Africa, Nigeria, and Uganda, for both primary and secondary levels). The individual student-level data generated by standard Koloso use is considerably richer than competition data, and more directly useful to teachers for adapting their instruction.

The Mini CEO competition data is therefore a specific, constrained dataset, but one that is useful precisely because it involves 107 publicly available, independent school-level observations, which is large enough to allow meaningful distributional analysis.

2. The observation: a near-normal distribution

The chart below shows the distribution of scores achieved by each of the 107 participating schools across the three Mini CEO games. The x-axis shows score bands from 0–10; the y-axis shows the number of schools achieving each score.

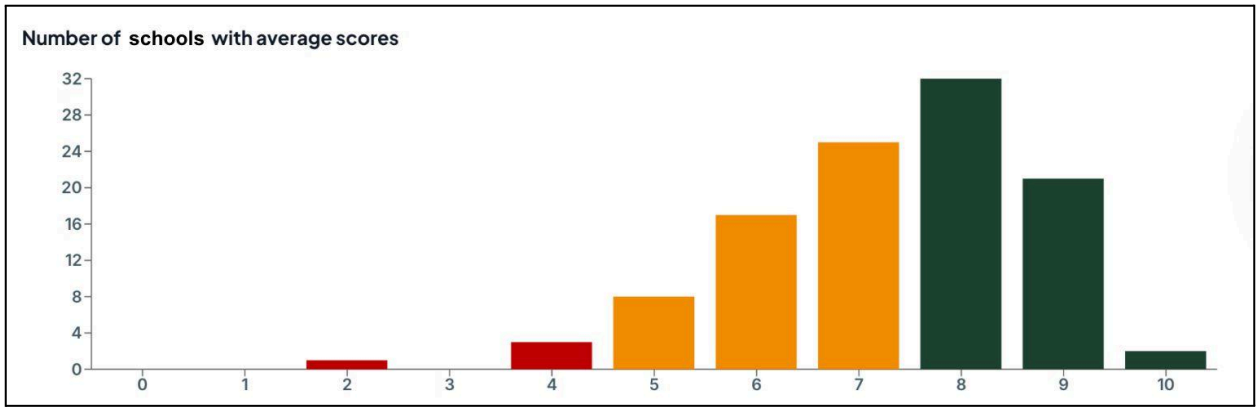

Figure 1: Distribution of school scores — Mini CEO Round 2 (n = 107 schools)

x-axis: Score band (0–10) · y-axis: Number of schools · Red = below 4, Orange = 4–6, Dark green = 7+

The distribution shows a clear bell-curve shape, peaking at score band 8 (approximately 31 schools), with secondary concentrations at 7 (25 schools) and 9 (21 schools). The tails are appropriately short: only 4 schools scored 4 or below, and only 2 schools achieved the maximum score of 10.

This is a notably clean result. An assessment that was poorly designed — too easy, too hard, or measuring nothing consistent — would produce either a flat distribution, a spike at one end, or an erratic pattern. A bell-curve of this shape, from 107 independent school observations, is the signature of an instrument that is discriminating genuine variation between schools.

3. What this result is testing

The implicit hypothesis

If Koloso's assessment questions are well-calibrated, meaning appropriately difficult, genuinely discriminating, and measuring real underlying variation in performance, then a sufficiently large sample of independent school teams should produce a score distribution that is approximately normal. Conversely, a poorly calibrated instrument would produce a flat, skewed, or erratic distribution.

The Round 2 result is consistent with this hypothesis. This is meaningful. It is not, however, proof of validity in the full psychometric sense - this would require demonstrating that the instrument measures what it claims to measure, not merely that it discriminates between schools. The distinction between discrimination and validity matters, and we return to it below.

4. What the evidence does and does not support

We can say:	The assessment is discriminating, 107 independent schools produced a spread of scores across the full range, concentrating in the middle. This is not trivial, suggesting that the questions are doing real work.
	The questions are appropriately calibrated for this cohort — not so easy that everyone scores 10, not so hard that scores collapse. The short tails in both directions confirm this.
	A near-normal distribution across 107 schools is a meaningful positive signal for the assessment methodology. This is a proof of concept - the first time Koloso has had enough publicly available independent data points to make this observation.
We cannot yet say:	The instrument is valid in the strict psychometric sense — that it reliably measures the specific learning competencies it claims to assess. This requires external validation against a recognised standard.
	The results are representative of all schools. The participating schools are self-selected (competition entrants), with a skew towards the private sector. Self-selected samples tend to over-represent more engaged and higher-performing institutions.
	This distribution would be replicated across a broader school population. The peak at 8 rather than 5 reflects the self-selection effect. A representative sample would likely show a lower mean and wider spread.
We can reasonably infer:	If curriculum-aligned digital assessment at this level of calibration were applied consistently across a representative, large sample of schools — say 500+ in a single country — the aggregate data would begin to function as a credible real-time picture of national learning performance.

5. The larger question: Can digital assessment complement or replace large-scale testing?

Large-scale periodic learning assessments (e.g. SACMEQ, EGRA/EGMA, PASEC, and national-level examinations) are the primary instruments through which education systems in sub-Saharan Africa currently generate data on learning outcomes. They are essential, but they are also enormously expensive (possibly in the range of US$0.5-1m, every 3-4 years for a country like Zambia), logistically complex, and structurally delayed (survey cycle typically 18-24 months) - by the time results are published and acted upon, the cohort assessed has often moved on by two or more grade levels.

The Global Partnership for Education's ANLAS initiative (Analysis of National Learning Assessment Systems) has documented these limitations across multiple country contexts, noting that the integration of classroom-level assessment data into national Education Management Information Systems is an area requiring systematic improvement in most partner countries. Zambia is currently in the process of strengthening its national learning assessment architecture, with a new national assessment programme being launched under UNICEF/FCDO support in 2026.

Koloso's assessment approach (curriculum-aligned, digitally administered, individually scored, and continuously available) is designed to produce exactly the kind of frequent, granular learning data that large-scale assessments cannot provide. The question is whether it can be shown to measure the same underlying constructs that established standardised instruments measure. If it can, the potential is significant: real-time, grade-by-grade, topic-by-topic learning data, updated daily, accessible to teachers, school leaders, district offices, and national planners simultaneously at a fraction of the cost of periodic survey-based assessments.

The Mini CEO Round 2 distribution is the first data point that makes this vision empirically plausible rather than merely theoretical. It does not prove the case. But it is consistent with it.

6. A proposed next step: a low-cost correlation study

6.1 The opportunity

Among the 107 schools that participated in Mini CEO Round 2, a subset are government schools that do not currently subscribe to Koloso. This creates a specific and time-limited opportunity. Zambia's UNICEF/FCDO national learning assessment programme is scheduled to assess learning outcomes in primary mathematics in 2026. If Koloso were to offer free platform access to the government school participants from the Mini CEO competition — for the remainder of the 2026 academic year — those schools would generate Koloso assessment data running in parallel with the national assessment.

This parallel dataset would allow a correlation analysis: do Koloso scores in curriculum-aligned mathematics assessment move together with scores on the UNICEF national assessment instrument in the same schools? A positive correlation would be strong evidence of construct validity, indicating that Koloso is measuring the same underlying learning as the national standard. A weak or absent correlation would be equally valuable, as it would identify where the methodology needs to improve.

6.2 Why this is the right next step

- **It is low cost to administer**: Free access to a defined subset of schools for the remainder of one academic year is a manageable investment with potentially high evidential return.
- **The comparison instrument is credible**: A UNICEF/FCDO national assessment is exactly the kind of established standardised instrument against which Koloso's validity should be tested.
- **The timing is right**: The national assessment is happening now. This window will not recur without deliberate effort.

- **It turns a limitation into a strength**: The fact that Mini CEO participants include non-subscribing government schools (usually seen as a sampling problem) becomes the source of an independent comparison group.
- **It is transparent and honest**: Our commitment to publishing the results (whatever they show) is consistent with Koloso's emphasis on evidence and would strengthen, not undermine, our credibility with UNICEF, GPE, and potential funders.

6.3 What we cannot control — and what we can

Koloso does not control the design of the Mini CEO competition, and Rounds 3 and 4 will proceed under the existing format. This is not a problem, and each additional round generates more data points that can be incorporated into any longitudinal analysis. We are also working with the Mini CEO organisers on the design of next year's programme, where (with support from sponsors) we hope that we will be able to incorporate a larger government school sample and a longer assessment window.

Independently of Mini CEO, Koloso can continue to improve the calibration and coverage of its own formative assessment question bank (e.g. expanding the number of questions per sub-topic, refining difficulty gradients, and extending coverage across all primary and secondary levels). This ongoing product development work is the foundation on which any future validity claim will rest.

7. Summary and proposed actions

What we have	107 schools, 3 games, near-normal score distribution, offering a meaningful proof of concept for Koloso's assessment methodology
What it shows	The assessment is discriminating and appropriately calibrated for this cohort
What it does not yet show	Construct validity, representativeness, or a causal link to learning outcomes
Immediate opportunity	Offer free Koloso access to government school participants to generate data parallel to the UNICEF/FCDO national assessment — enabling a correlation study
Medium-term actions	Incorporate Round 3 and 4 data; improve question bank calibration and coverage; publish results transparently
Longer-term vision	Curriculum-aligned digital assessment at scale as a real-time complement to — and eventual partial substitute for — periodic large-scale testing

Disclosure and limitations

This working paper is authored by the CEO of Koloso Ltd and describes evidence relating to Koloso's own product. Readers should apply appropriate scrutiny to the findings and conclusions. Koloso is committed to transparent evidence-building and will publish the results of future validation studies regardless of outcome. The Mini CEO score data referenced in this paper will be made available on request to accredited researchers.