**KOLOSO** Working Paper WP/2026/02
*Evidence & Impact Series*

# Continuous curriculum-aligned digital assessment: Evidence foundations and a roadmap for building the evidence base in live school deployment

*Can continuous, curriculum-aligned digital assessment generate reliable instructional and system-level learning signals at low cost? Evidence foundations, analytical potential, and a practical strategy for strengthening the evidence base as Koloso scales in schools.*

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

**Abstract**

The Koloso Assess module is designed to generate continuous, curriculum-aligned digital assessment signals within the teaching process. Unlike traditional periodic tests or large-scale learning assessments, Koloso assessments are short, frequent, and embedded directly in classroom practice across both primary and secondary levels.

This working paper (the second in Koloso's Evidence and Impact Series) provides the theoretical foundations for that approach. WP/2026/01 demonstrated that Koloso's assessment instrument is discriminating and calibrated, producing a near-normal distribution of scores across 109 independent school teams. WP/2026/02 explains why that result matters: it reviews the research on formative assessment, retrieval practice, spaced learning, and low-stakes testing that underpins Koloso's assessment design; describes the types of learning signals the system is designed to generate; and sets out the limits of what can currently be claimed.

Finally, the paper proposes a pragmatic strategy for strengthening the evidence base as Koloso is deployed at scale - prioritising low-cost, high-value evidence through correlation studies, longitudinal platform data analysis, and transparent publication of findings.

## 1. Background and context

This paper builds directly on WP/2026/01, which examined the distribution of scores across 109 schools in the Mini CEO Round 2 competition and concluded that Koloso's assessment instrument is discriminating genuine variation between schools. The resulting near-normal bell-curve distribution is a proof of concept: evidence that the underlying measurement tool is functioning as designed. WP/2026/02 provides the theoretical framework that explains what the tool is designed to do, why it is designed that way, and what systematic deployment will allow us to learn.

### 1.1 The role of assessment in Koloso's teaching loop

The Koloso Teaching Support System organises teaching around a structured five-stage loop: Plan → Teach → Assess → Analyse → Adapt. The Assess module occupies a specific and deliberately bounded role in this architecture: it generates the assessment instruments that produce evidence, and captures the resulting data. It does not interpret that data — interpretation is the function of the Analyse module.

This separation between measurement and interpretation is deliberate. It maintains conceptual clarity between the evidence generated and the diagnostic insights derived from it, and ensures that teachers engage actively with the analysis process rather than simply receiving pre-packaged conclusions.

### 1.2 How Koloso's assessment model differs from traditional approaches

Traditional classroom assessment is episodic: topic tests, mid-term tests, and end-of-term examinations occurring at fixed points in the teaching calendar. Results arrive after teaching decisions have already been made, limiting their instructional usefulness.

Koloso's model is designed around continuous low-stakes assessment embedded within teaching. Most Koloso assessment formats involve short quizzes of around 10 questions, completed within approximately two minutes, and repeated frequently across the teaching cycle. The aim is to generate lightweight, actionable signals about student understanding without interrupting instructional flow — signals that, when aggregated across lessons and topics, reveal learning patterns that are invisible in periodic testing alone.

## 2. Evidence foundations

Koloso's assessment approach draws on several well-established strands of educational research, each of which has particular relevance to the primary and secondary school contexts in which Koloso operates.

### 2.1 Formative assessment and instructional feedback

The evidence base for formative assessment — assessment used to inform and adapt teaching rather than to grade students — is among the strongest in educational research. Black and Wiliam's landmark review (1998) of over 250 studies concluded that formative assessment practices can significantly improve learning outcomes when teachers adapt instruction in response to assessment evidence. Hattie's synthesis of meta-analyses (2009) identified feedback as one of the most powerful influences on student achievement.

The key factor in both bodies of research is timeliness: feedback must arrive while teaching decisions can still be adjusted. Lesson-level and topic-level assessments that return results within the same teaching cycle — as Koloso's do — are designed to meet this condition.

### 2.2 Retrieval practice and the testing effect

A second foundation is the 'testing effect' — the consistent experimental finding that retrieving knowledge from memory strengthens learning more effectively than passive review. Roediger and Karpicke (2006) demonstrated that students who repeatedly retrieve information retain it significantly better over time than students who reread the same material. Karpicke and Blunt (2011) confirmed that retrieval practice can outperform other study strategies for long-term retention.

Koloso operationalises retrieval practice through daily challenge quizzes, spaced retrieval sessions, spiral assessments, and cumulative revision series — each of which prompts active recall rather than passive recognition.

### 2.3 Spaced learning and retention

Memory research originating with Ebbinghaus demonstrates that forgetting follows a predictable curve unless learning is revisited at appropriate intervals. Distributed practice — returning to material after increasing time gaps — significantly improves long-term retention (Cepeda et al., 2006). Koloso's spiral assessment formats and structured revision series are designed around this principle, revisiting earlier topics before knowledge decays beyond retrieval threshold.

### 2.4 Low-stakes testing and motivation

Frequent low-stakes testing can improve both retention and engagement. Studies show that regular quizzes enhance learning outcomes without the negative motivational effects associated with high-stakes testing (Bangert-Drowns, Kulik and Kulik, 1991; Roediger et al., 2011). Because Koloso quizzes are short, repeatable, and formative — designed as learning tools rather than judgement instruments — they are consistent with this evidence.

### 2.5 The African classroom context

The research literature summarised above is predominantly drawn from Western school contexts. However, structured, curriculum-aligned approaches to formative assessment have a specific and growing evidence base in sub-Saharan Africa. RTI International's evaluations of structured pedagogy programmes in Kenya, Uganda, and Ethiopia have found that curriculum-aligned instructional materials and formative assessment practice can improve student learning outcomes by 0.3–0.6 standard deviations without changes to class size, teacher qualifications, or infrastructure (RTI International, various).

This evidence base is directly relevant to Koloso's context: large class sizes, variable teacher training, limited resources, and national curriculum frameworks that differ significantly from those in higher-income countries. Koloso's assessment design reflects these realities — questions are calibrated for the specific curricula of Zambia, South Africa, Nigeria, and Uganda; language is kept accessible for students whose instruction may be in a second language; and the platform is designed to function in low-connectivity environments.

## 3. What Koloso assessments are designed to measure

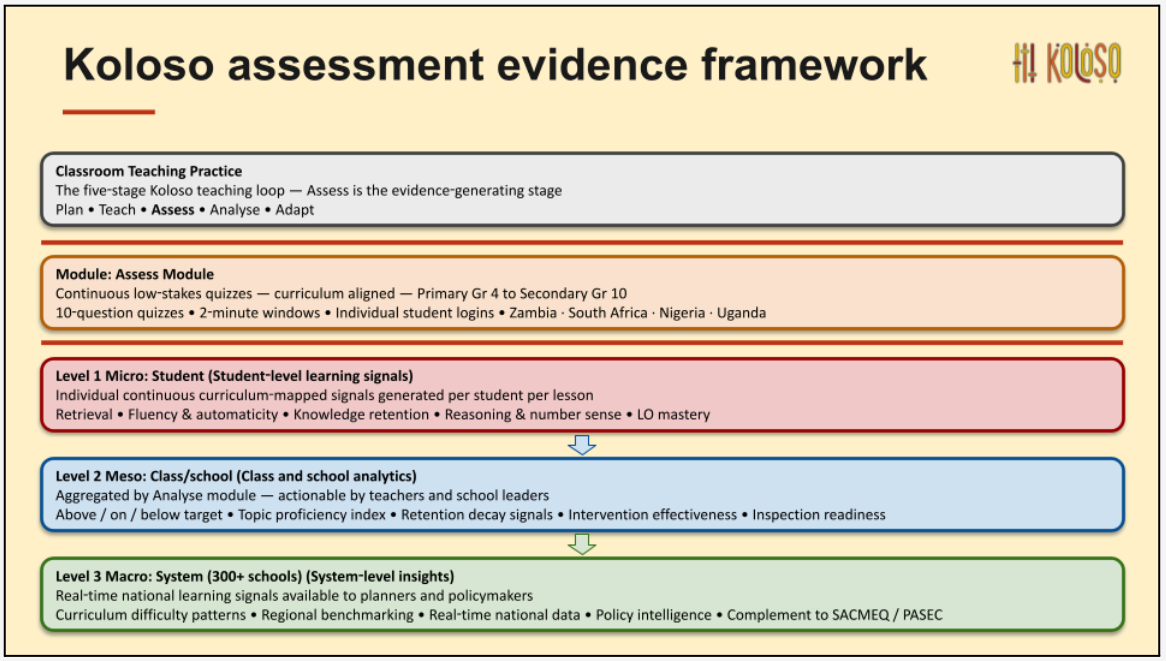
Koloso assessments operate across three analytical layers simultaneously, generating different types of learning signal depending on the format and timing of the assessment.

| Assessment layer | What it measures | What it tells the teacher |
|---|---|---|
| **Instructional layer(immediate feedback)** | Whether students understood material just taught — lesson-level and learning-objective diagnostics | Which students are ready to move on; which need re-teaching before the next lesson; which misconceptions are common across the class |
| **Memory layer(retention signals)** | Whether earlier learning has been retained — spaced retrieval assessments, spiral checks, fluency and automaticity measures | Which topics are consolidating in long-term memory; where forgetting is occurring; which students need additional retrieval practice |
| **Certification layer(summative signals)** | End-of-topic and end-of-term performance levels — benchmarks, mock examinations, formal topic assessments | Where each student stands relative to curriculum expectations; which cohorts are on track for examinations; school-level performance against targets |

In standard Koloso deployment, every student has a unique login and receives an individualised sequence of assessments that are grade-specific and curriculum-aligned to their national syllabus. The data generated at the instructional and memory layers is therefore student-level, granular, and continuous — in contrast to the school-level, aggregate, and periodic data generated by large-scale assessments.

Figure 1 illustrates how assessment data generated by the Assess module flows into the Analyse module and then through three levels of aggregation — from individual student signals to class-level patterns to system-level insights.

**Figure 1: Koloso Assessment Evidence Architecture — from student signal to system insight**




Classroom teaching (Plan → Teach → Assess → Analyse → Adapt) → Student-level learning signals → Class / school analysis → System-level insights

## 4. What the system allows us to analyse

The continuous individual assessment data generated by standard Koloso deployment enables several types of analysis that are not possible with periodic testing alone.

### 4.1 Within the classroom: Instructional decisions

At the lesson and topic level, aggregated student results allow teachers to identify common misconceptions, distinguish students who are above, on, or below target for each topic, and evaluate whether a teaching adaptation has been effective. The performance category dashboard shown in WP/2026/01 (Figure 2) illustrates this capability at school-team level; in standard deployment it operates at individual student level, updated continuously.

Key analytical signals produced by the Analyse module for classroom use include learning objective mastery rates (individual and class-level, with trend over time), topic proficiency distribution (proportion of students above, on, or below target), retention decay indicators (changes in accuracy between assessments, time since last exposure), fluency signals (response accuracy under time pressure), and intervention effectiveness tracking (score improvement after targeted re-teaching).

### 4.2 Across the school: School-leader intelligence

Aggregated across classes and grade levels, Koloso data provides school leaders with a continuous picture of teaching quality and curriculum coverage. This is directly relevant to school inspection readiness in all four of Koloso's current markets, where inspection frameworks include expectations about

evidence of formative assessment practice and teacher file documentation. A school leader who can show continuous assessment data, topic mastery trends, and evidence of teaching adaptation is better prepared for inspection than one who can only produce periodic test scores.

### 4.3 At scale: System-level learning signals

At sufficiently large deployment — hundreds of schools across a country — the aggregate data generated by Koloso begins to function as a real-time national learning signal. Grade-by-grade, topic-by-topic mastery rates, updated continuously, would provide education planners with information that no periodic survey can offer: not a snapshot taken every three to four years, but a continuously updated picture of where students are struggling across the curriculum and where teaching is working.

This is the larger vision articulated in WP/2026/01. The empirical validation of that vision — whether Koloso scores at scale correlate with what established national assessment instruments measure — is the subject of the proposed correlation study described in Section 7 below.

## 5. What the system cannot measure

Intellectual honesty requires clarity about the limits of digital multiple-choice assessment. Koloso can assess competencies that are measurable through structured questions with discrete correct answers. It cannot currently capture:

- Extended written reasoning and mathematical proof
- Practical or laboratory work
- Oral explanation, discussion, and argumentation
- Collaborative problem solving
- Creative and open-ended tasks

These dimensions of learning remain within the domain of teacher observation, peer assessment, and portfolio-based approaches. Koloso is designed to complement teacher professional judgement, not replace it.

Koloso also does not yet implement full item-response theory (IRT) calibration at the scale used in national standardised assessments. The system uses statistical calibration to balance question difficulty, but the rigorous psychometric standardisation applied to SACMEQ or PASEC instruments involves large-scale field-testing and equating procedures that are beyond Koloso's current resources. This is a known limitation and a priority for future development.

Finally, observational platform data cannot establish causal claims about learning improvement. Showing that students who use Koloso more frequently achieve higher scores does not prove that Koloso caused the improvement — other explanations, including pre-existing school quality differences, are possible. Establishing causal claims requires experimental or quasi-experimental study designs, which are described in the research agenda below.

## 6. Research questions for empirical investigation

As Koloso is deployed at scale, the platform will generate large volumes of assessment data across lessons, topics, schools, and countries. This creates concrete opportunities to investigate questions that matter both for Koloso's product development and for the wider evidence base on digitally mediated formative assessment in African school contexts.

WP/2026/01 addressed the first of the questions below in preliminary form: the near-normal distribution of Mini CEO Round 2 scores is an early signal relevant to predictive validity. The following agenda sets out what systematic investigation of deployment data will allow us to examine more rigorously.

### 6.1 Predictive validity of continuous assessment signals

Research question: *Do early Koloso assessment signals predict later performance in formal assessments?*

If lesson-level or topic-level assessment results correlate with later examination or national assessment performance, this would suggest that Koloso provides reliable early indicators of learning progress — and would substantially strengthen the case for its use as a complement to periodic large-scale testing.

Possible analysis: correlation between Koloso topic mastery scores and end-of-term test results; prediction of national assessment performance using earlier Koloso data; identification of early warning indicators for learning difficulties.

### 6.2 Impact of retrieval practice on knowledge retention

Research question: *Does frequent retrieval practice in Koloso improve long-term retention compared with standard teaching practice?*

Koloso incorporates retrieval events through daily challenges, spaced quizzes, and spiral checks. Deployment data provides a natural opportunity to examine whether higher retrieval frequency is associated with stronger retention outcomes within the same class or across comparable classes.

Possible analysis: comparison of retention scores between students with high and low retrieval practice frequency; measurement of retention decay across different assessment schedules; evaluation of spiral formats as reinforcement mechanisms.

### 6.3 Effectiveness of instructional adaptation

Research question: *Do teaching adaptations informed by Koloso assessment results lead to measurable improvements in student learning?*

Formative assessment research claims that instruction improves when teachers adjust their teaching based on assessment evidence. Koloso's Adapt module creates a natural environment for examining this process — comparing student performance before and after targeted interventions, and testing whether classes where teachers adapt more frequently show stronger learning gains.

### 6.4 Learning trajectory patterns

Research question: *What patterns of learning progress emerge across different student groups over time?*

Continuous assessment allows learning to be observed as a trajectory rather than a snapshot. Platform data may reveal typical patterns — including early difficulty followed by recovery, or stable plateaux — and allow comparison of trajectories across demographic groups, school types, and national contexts.

### 6.5 System-level curriculum difficulty signals

Research question: *What curriculum areas consistently produce learning difficulties across schools?*

Aggregated data may reveal structural challenges in curriculum sequencing or concept difficulty that are invisible in school-level analysis. Identifying topics with consistently low mastery rates across a country, or transitions where performance reliably declines, would be of direct value to curriculum developers and education planners.

### 6.6 Relationship between assessment frequency and learning outcomes

Research question: *Does the frequency of low-stakes assessment influence learning progress?*

Koloso's design assumes that frequent, embedded assessment supports learning through retrieval and feedback. Deployment data allows direct comparison of learning gains between classes using assessments at different frequencies, testing this assumption with real-world data.

## 7. Building the evidence base: a practical roadmap

Koloso operates as a resource-constrained startup. Evidence must therefore be built incrementally, with each stage producing findings that inform the next. The following four-stage roadmap prioritises low-cost, high-value approaches that can be executed alongside — not instead of — product development and commercial scaling.

| | |
|---|---|
| **Stage 1: Observational evidence** | Analyse data from early deployments and competitions. Describe score distributions, topic difficulty patterns, and usage trends. Publish findings as working papers. (Underway: WP/2026/01 is Stage 1 output.) |
| **Stage 2: Correlation studies** | Compare Koloso assessment results with national assessment instruments in the same schools. The proposed government school pilot described in WP/2026/01 — running free Koloso access in parallel with the UNICEF/FCDO Zambia national assessment — is the first Stage 2 study. Target: Q2–Q3 2026. |
| **Stage 3: Longitudinal studies** | Use platform data from subscribing schools to track learning trajectories over one or more academic years. This requires a sufficient base of consistently-using schools (target: 50+ schools, 2 or more terms of data). Target: 2026–2027. |
| **Stage 4: Collaborative research** | Partner with universities, research organisations, or ministries of education for peer-reviewed, independently validated studies. This is the stage at which causal claims about learning improvement can be made with credibility. Target: 2027 onwards, dependent on Stage 2 and 3 results. |

**Commitment to transparent publication**

Koloso will publish the results of each stage in this series, regardless of outcome. Working papers that report weak correlations or null findings are as valuable as those that report positive results — they tell us where the methodology needs to improve and build the credibility of findings that do show positive effects.

## 8. Conclusion

Continuous curriculum-aligned digital assessment offers a distinctive and potentially valuable complement to traditional periodic testing — one that generates instructionally actionable signals at the moment they are most useful, and that can aggregate upward from individual student data to class, school, and national-level insights.

The approach is grounded in a well-established research base: formative assessment, retrieval practice, spaced learning, and low-stakes testing each have strong and consistent evidence of effectiveness. The growing evidence base from structured pedagogy programmes in sub-Saharan African contexts — including Kenya, Uganda, and Ethiopia — provides specific support for curriculum-aligned, assessment-informed approaches in the school contexts Koloso serves.

WP/2026/01 provided the first empirical signal that Koloso's assessment instrument is functioning as designed — discriminating genuine variation between schools and producing a calibrated score distribution. WP/2026/02 has set out why that signal matters: the theoretical framework that makes a discriminating instrument useful, and the research agenda that will determine whether it is also valid.

The most important next step is the proposed correlation study in WP/2026/01: offering free Koloso access to government school participants from the Mini CEO competition, running alongside the UNICEF/FCDO national learning assessment programme in Zambia. If Koloso scores in curriculum-aligned mathematics assessment move together with scores on the national assessment instrument in the same schools, that result would substantially advance the case for continuous digital assessment as a low-cost, real-time complement to periodic large-scale testing.

For a small company operating in one of the world's most challenging and most important educational contexts, this is the right kind of ambition: proportionate to our current resources, honest about our limitations, and oriented toward the evidence that would make a real difference to the schools and children we are trying to serve.

## References

Bangert-Drowns, R., Kulik, C., and Kulik, J. (1991). Effects of frequent classroom testing. Journal of Educational Research, 85(2), 89–99.

Black, P., and Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. King's College London.

Cepeda, N.J., et al. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological Bulletin, 132(3), 354–380.

Hattie, J. (2009). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge.

Karpicke, J., and Blunt, J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. Science, 331(6018), 772–775.

Roediger, H., and Karpicke, J. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249–255.

Roediger, H., Putnam, A., and Smith, M. (2011). Ten benefits of testing and their applications to educational practice. Psychology of Learning and Motivation, 55, 1–36.

RTI International. (various). Structured pedagogy evaluation reports: Kenya, Uganda, Ethiopia. RTI International / USAID.

**Disclosure and limitations**

This working paper is authored by the CEO of Koloso Ltd and discusses 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.

*Koloso Ltd | WP/2026/02 | Continuous Curriculum-Aligned Digital Assessment | March 2026 | koloso.app/evidence | james@koloso.app*