ETDP-SETA Phase 2

AI-Supported Evidence-Based Mathematics Pedagogy

A Five-Year Pilot Implementation

Free State Province, South Africa

Submitted by:	Mathematics ETDP-SETA Research Chair
Institution:	University of the Free State
Duration:	Five Years
Target:	10–15 Schools
Date:	September 19, 2025

Transforming Mathematics Education Through AI Innovation

Contents

1 Executive Summary

Project Overview

This proposal seeks ETDP-SETA Phase 2 support for a five-year AI-supported mathematics education pilot across 10–15 Schools in the Free State. The project strengthens foundational numeracy through real-time diagnostic assessment, personalized learning support, and data-informed pedagogy, while rigorously evaluating impact, equity, cost-effectiveness, and scalability. Partnering with the Free State Department of Education, universities, and pilot schools, we aim to demonstrate a viable, sustainable model for provincial and national scale.

1.1 Problem, Opportunity, and Aim

Persistent underperformance in mathematics and large equity gaps constrain the STEM pipeline. Advances in AI-enabled diagnostics and teacher dashboards make it feasible to identify misconceptions early and personalise instruction—also in resource-constrained contexts. Over Five Years, we will deploy, support, and evaluate an adaptive pedagogy model across 10–15 Schools, with explicit focus on impact, equity, cost-effectiveness, and feasibility for provincial scale.

1.2 Objectives and Outcomes

Over five years, we will raise standardized mathematics performance in pilot schools by approximately 0.25 standard deviations. We will halve achievement gaps between learners in Q1–Q3 and Q4–Q5 schools and lift teacher data-literacy and formative assessment practice to above 4/5 proficiency. In parallel, we will establish durable provincial implementation capacity and a policy-tested pathway for scaling the model beyond the initial cohort of schools.

1.3 Approach and Design

The intervention integrates open, offline-capable AI diagnostics, learner-facing feedback, teacher dashboards, targeted content, and coaching. We adopt a mixed-methods design anchored by a stepped-wedge cluster randomized trial (CRT): all pilot schools are randomly assigned to one of three activation waves on pre-specified dates. This yields within- and between-school contrasts while ensuring universal participation. Formative learning continues alongside (without altering the randomized schedule), and independent summative evaluations are embedded in Years 3 and 5. Implementation follows a phased rollout from three to fifteen schools with continuous professional development, local technical support, and strong district partnerships.

1.4 Value for Money

The total investment of R 45 750 000 over five years finances technology, training, embedded coaching, local technical support, and rigorous independent evaluation.

Unit costs (transparent):

- *Per learner-year:* R 8 632 based on 5 300 learner-years (300,500,1000,1500,2000).
- *Per unique learner over five years:* approximately R 22 875 at steady-state reach of 2,000 learners.

1.4.1 Cost-Effectiveness: Worked Example

METHOD: Rand per 0.1 SD = *Total economic cost* ÷ (*aggregate SD gain* / 0.1), where aggregate SD gain = learners × expected mean gain (SD).
Base-case assumptions:

- Annual costs use the budget by year (R 12.10m, 10.00m, 8.80m, 7.95m, 6.90m).
- Expected mean gain per treated learner-year: 0.15 SD (Y2), 0.20 SD (Y3), 0.25 SD (Y4–Y5). Y1 is setup/baseline (no effect).
- *Dosage condition*: effects assume minimum viable exposure*.
- Sensitivity range shows simultaneous ±20% on costs and effects (best: −20% cost, +20% effect; worst: +20% cost, −20% effect).

Year	**Total cost** (R m)	**Learners** (n)	**Mean gain** (SD)	**Aggregate gain** (learner-SD)	**R per 0.1 SD** base [best–worst]
Y1 (baseline)	12.10	300	–	–	N/A
Y2	10.00	500	0.15	75	R 13,333 [8,889–20,000]
Y3	8.80	1,000	0.20	200	R 4,400 [2,933–6,600]
Y4	7.95	1,500	0.25	375	R 2,120 [1,413–3,180]
Y5 (steady state)	6.90	2,000	0.25	500	R 1,380 [920–2,070]

** Dosage: protected timetable with ≥60 minutes/week on-platform practice across 30 instructional weeks; ≥3 diagnostic cycles/year; teacher dashboard weekly active rate ≥80%; feedback-driven grouping in the protected period. Effects are contingent on meeting these fidelity gates; shortfalls reduce realized gains proportionally.*

Value for money is further enhanced through open-source components, shared infrastructure, and structured capacity transfer to provincial teams by Year 5.

1.5 Expected Impact

We expect measurable gains in learner attainment, reduced inequities in achievement, and sustained improvements in teacher practice through data-informed instruction and coaching. The project will leave behind institutional capability, evidence, and implementation tools to enable expansion at provincial and national levels.

1.6 Strategic Alignment

This comprehensive proposal presents a transformative initiative aligned with:

- ✓ **National Development Plan 2030** - Quality basic education objectives
- ✓ **ETDP-SETA Strategic Plan** - Skills development in education sector
- ✓ **Department of Basic Education** - Mathematics improvement priorities
- ✓ **4IR Strategy** - Digital transformation in education

1.7 Key Innovation Features

AI-Powered Diagnostics: Real-time assessment and personalized learning pathways for each learner

Evidence-Based Approach: Continuous monitoring and iterative improvement

Teacher Empowerment: Data-driven insights and professional development support

Scalable Solution: Open-source platform designed for provincial and national expansion

2 Context and Rationale

2.1 The Mathematics Crisis in South Africa

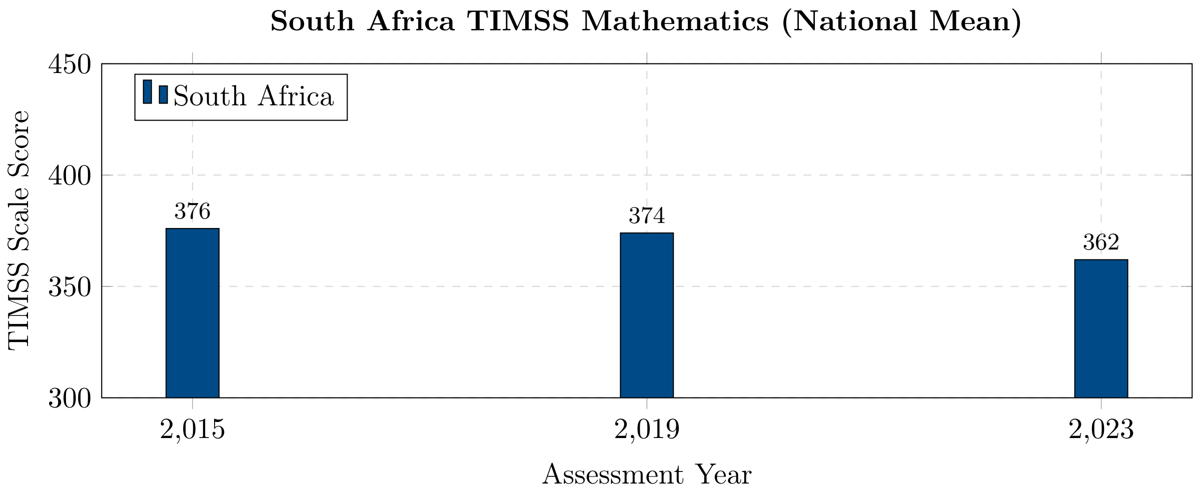


Figure 1: South Africa TIMSS mathematics mean scores (2015, 2019, 2023). Source: IEA (2023) TIMSS International Results in Mathematics; national means reported. Provincial TIMSS results are not published.

2.2 Hardware Itemization and Assumptions

Device / Service	**Unit cost (R)**	**Qty / school**	**Qty (15 schools)**	**Notes**
Android tablets (10 in, rugged case)	3 800	66	990	60 devices + 10% spares per school
Protective cases (drop/shock)	250	66	990	*Already included in tablet unit cost; not costed separately*
Charging cart / lockable trolley	8 500	1	15	Per school; secure storage and charging
AP + router kit (dual WAN)	3 000	1	15	Local caching; offline sync support
UPS (line-interactive, 1 kVA)	5 000	1	15	Graceful shutdown; mitigates brief outages
MDM license (per device-year)	120	–	990 × 5	Five-year license for enrolled fleet
Replacement allowance (Y3–Y5)	3 800	–	238	8%/year attrition across three years (tablets)
Subtotal (hardware+MDM, 5y)				˜*R 5.8 m* (ties to budget "Hardware and devices")

Assumptions: 66 devices per school (60 active, 10% spares) support protected periods (small-group rotation). MDM provides fleet enrollment, remote lock/wipe, and usage policy enforcement; caching reduces bandwidth dependence. Replacement allowance covers loss/damage from Y3 onward; procurement uses provincial transversal rates where available.

2.3 Coaching Delivery Model

Delivery approach. Cluster coaching with on-site, small-group sessions blended with classroom-embedded cycles. At steady state the operating headcount is *one coach per three schools.* Coaches run bi-weekly 45-minute small-group coaching per teacher, classroom observations and feedback in protected periods, one lesson-study cycle per term, and two termly cluster workshops. This aligns with the dosage specified in Teacher Professional Development Dosage.

Coach-to-teacher ratio. The coached cohort is the mathematics teaching team per school (Foundation Phase plus HOD), planned at ≈10 teachers/school. Ratios therefore evolve with rollout and operating headcount:

Year	Schools active	Teachers coached	Coach headcount	Ratio (teachers:coach)
Y2	5	≈50	3	≈ 17:1
Y3	10	≈100	4	≈ 25:1
Y4	15	≈150	5	≈ 30:1
Y5	15	≈150	5	≈ 30:1

Schedule. Each coach allocates two on-site days per school per month (6 days/month at steady state), with weekly remote check-ins (15–20 min) and termly cluster workshops (one half-day).

Classroom-embedded cycles occur in the protected period; observation and feedback notes are logged to the coaching tracker for moderation.

Travel and logistics. Routing groups neighbouring schools to a single coach minimises travel time. Typical travel is ≤1,000 km/month/coach within district clusters; travel time is budgeted as 15% of coach time. The budget covers mileage or shuttle, occasional overnight stays for remote sites, per diem, and shared facilitation materials. Coaches carry a spare device kit (2 tablets, charger block) for on-site contingencies.

Staffing and cost reconciliation. Coaching delivery is funded under Human Resources ("Coaching delivery"). Costs reconcile to the budgeted R 5,000,000 over five years with an average loaded cost of R 333,333 per FTE-year (salary, benefits, travel, materials). Operating headcount varies across the year with term breaks; the FTE-years are therefore lower than headcount. The planned FTE-years and costs are:

Year	Coach headcount (operating)	FTE-years	Cost per FTE-year	Annual cost
Y1 (setup, baselines)	2	1.5	R 333,333	R 500,000
Y2	3	3.0	R 333,333	R 1,000,000
Y3	4	4.0	R 333,333	R 1,333,333
Y4	5	4.0	R 333,333	R 1,333,333
Y5 (handover)	5	2.5	R 333,333	R 833,333
Total		*15.0*		*R 5,000,000*

The staffing plan keeps the effective ratio at or below 30:1 once all 15 schools are active while staying within the HR coaching budget. During the Year 5 handover, coaches focus more on mentoring district facilitators, explaining the lower FTE while maintaining operating headcount during terms.

2.4 Free State Provincial Context

Indicator	**Free State**	**National**
Quintile 1–3 schools (no-fee)	68%	60%
Teacher vacancy rate	12%	10%
Learner-teacher ratio	32:1	30:1
Digital infrastructure availability	23%	31%
Grade 3 numeracy achievement	41%[a]	47%[a]

Table 1: Key education indicators highlighting intervention need

Notes: Teacher data-literacy score uses rubric **TL-DL v1.1** (termly moderation); Learner engagement index from platform analytics (weekly aggregate, termly reported); Pass rates from provincial assessments (annual). Platform SLO measured monthly with external uptime monitor. *a) Grade 3 numeracy achievement reflects the pass rate (≥ minimum pass threshold) from DBE National Assessment Results 2023 for Free State and national aggregates. The 57% figure cited below refers to the proportion meeting the provincial minimum mathematics competency benchmark in the 2019 Free State Systemic Evaluation; thresholds and instruments differ and are not directly comparable (DBE & Free State DoE, 2019). Unless otherwise stated, baseline indicators are sourced from DBE (2023).*

2.5 Provincial Baseline and Evidence Summary

In 2019, approximately 57%[b] of Free State Grade 3 learners achieved the provincial minimum mathematics competency benchmark, signalling substantial foundational gaps by the end of the Foundation Phase (DBE & Free State DoE, 2019). This early shortfall compounds in later grades, widening achievement disparities and reducing participation and success in the STEM pipeline. South Africa's mathematics education quality is also reported among the lowest globally, underscoring the urgency of targeted, evidence-based interventions (ETDP-SETA Landscape Review, 2025; OECD, 2020).

b) 2019 Free State Systemic Evaluation; competency benchmark differs from DBE 2023 pass-rate definition.

Current assessment practice relies heavily on infrequent summative examinations, which reveal problems only after misconceptions have become entrenched. Teachers often lack timely, granular diagnostic information to guide responsive remediation, particularly in resource-constrained schools. Addressing these constraints requires a system that delivers continuous, classroom-embedded evidence and supports teachers to act on it.

2.6 Evidence Base for AI-Supported Pedagogy

Evidence from intelligent tutoring and adaptive learning systems shows measurable gains in mathematics achievement relative to traditional instruction (e.g., Ma et al., 2014; Cheung & Slavin, 2013; OECD, 2020). AI-enabled diagnostics can raise learner engagement and reduce teacher workload by automating feedback and surfacing misconceptions for targeted intervention (Woolf et al., 2009). Successful adoption, however, requires structured teacher professional development, strengthened data-use skills, and safeguards for privacy, fairness, and transparency, in line with POPIA and national research ethics guidance (ASSAf, 2018; ASSAf, 2023).

3 Methodology

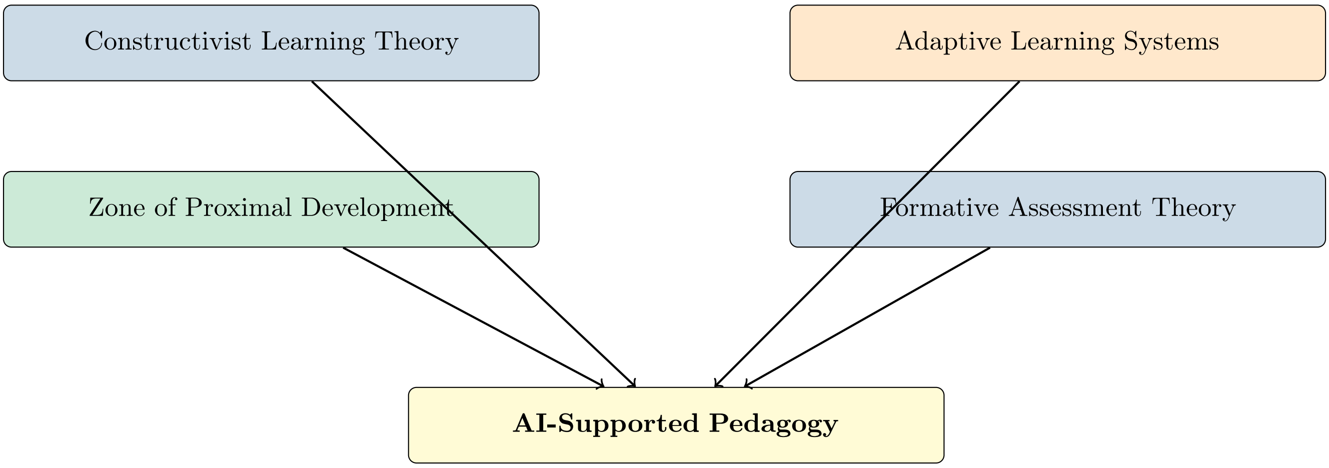
3.1 Research Design

We implement a **stepped-wedge cluster randomized design** across participating schools, introducing the intervention in three waves at fixed intervals. This yields within-school and between-school contrasts while ensuring all sites ultimately receive the program. Quantitative analysis follows an intention-to-treat (ITT) multilevel model with school random effects and cluster-robust intervals; qualitative components (case studies, observations, interviews) explain mechanisms of change; a developmental evaluation lens supports iteration throughout the five years.

Mixed-Methods Approach

Quantitative: Stepped-wedge cluster randomized rollout; subgroup and equity analyses; cost-effectiveness.
Qualitative: Case studies, classroom observations, teacher interviews, and focus groups.
Developmental: Continuous improvement cycles with rapid feedback to implementers.

3.2 Theoretical Framework


Constructivist Learning Theory
Adaptive Learning Systems
Zone of Proximal Development
Formative Assessment Theory
AI-Supported Pedagogy

3.3 Measurement and Item Bank

Learner achievement will be measured with an item bank aligned to CAPS strands and cognitive demand levels. Items will be authored and piloted annually and calibrated using 2PL/3PL Item Response Theory models. Diagnostic items include misconception-targeted distractors mapped to a skills graph; annual anchor sets are protected from classroom exposure and rotated across forms to preserve equating validity. Anchor items will support equating across years, enabling comparable scale scores over time. We will conduct differential item functioning (DIF) analyses across language and quintile to monitor fairness, remove or revise biased items, and maintain a transparent *skills map* that links items to knowledge components and prerequisites for adaptive sequencing. Items will be available in isiZulu, Sesotho, Afrikaans, and English with rigorous translation, back-translation, and proofing; DIF remediation will include linguistic review in addition to statistical screening.

3.4 Research Objectives and Questions

The study will quantify the effect of AI-supported, data-driven pedagogy on learner mathematics outcomes over five years; strengthen teacher data-literacy and formative assessment practice through dashboards and coaching; assess equity effects across school quintiles (Q1–Q5), gender, and language groups; and determine cost-effectiveness and implementation feasibility for provincial scale.

We ask: What is the impact of the intervention on standardized mathematics scores relative to *not-yet-treated periods* in a stepped-wedge CRT; through what teacher-practice mechanisms (diagnostic use, grouping, feedback) are gains achieved; how do effects vary by learner subgroup and school context; what is the cost per 0.1 SD gain and the projected cost of scale; and what fidelity, acceptability, and feasibility conditions are required for sustained adoption?

3.5 Hypotheses

We hypothesize that learners in intervention schools will achieve at least a 0.25 standard deviation improvement over five years relative to *control periods* in the stepped-wedge schedule; that teachers using dashboards and coaching will show significant gains in data-use and formative assessment quality; that equity gaps will narrow more during treated periods than during control periods;

and that the intervention will be cost-effective relative to comparable mathematics improvement programs in South Africa.

4 Implementation Plan

Implementation will proceed in phases that balance product development, capacity-building, and rigorous evaluation. Early years focus on establishing baseline measures, co-designing the platform with teachers, and validating diagnostic instruments in three to five schools. Subsequent years expand the rollout, deepen professional development, and stress-test the support model across diverse contexts. The rollout follows a **stepped-wedge cluster randomized** schedule in three waves so that all sites ultimately participate while allowing credible impact estimation. Throughout, evidence from monitoring and formative evaluation will inform iterative improvements, culminating in an external summative evaluation and a sustainability and scale roadmap by Year 5.

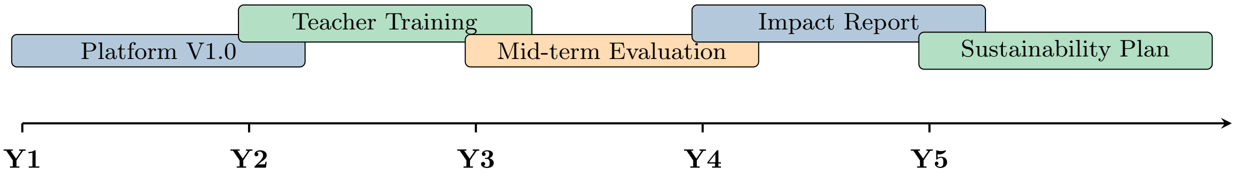
4.1 Teacher Professional Development Dosage

Teachers receive bi-weekly 45-minute coaching sessions focused on data use and responsive instruction, two termly cluster workshops for collaborative planning and dashboard practice, and one exemplar lesson-study cycle per term. District and school leadership will provide release time or relief arrangements to protect coaching and workshop participation without increasing overall workload.

4.2 Five-Year Phased Approach

Phase	Key Activities	Schools	Learners
Year 1	Platform development, baseline studies	3	300
Year 2	Pilot implementation, iterative refinement	5	500
Year 3	Expanded rollout, teacher training	10	1,000
Year 4	Full implementation, impact assessment	15	1,500
Year 5	Sustainability planning, scale preparation	15	2,000

4.3 Key Deliverables Timeline



4.4 Stepped-Wedge Schedule

	Wave 1	Wave 2	Wave 3
Activation (term)	Y2 T1	Y3 T1	Y4 T1
Schools (approx.)	5	5	5

Contamination controls include access scoping per school, no cross-school sharing of credentials before crossover, and phased enablement of features aligned to the schedule.

5 Expert-Informed Project Costing

5.1 Total Budget Overview

Total Project Investment: R 45,750,000
(Five-year implementation period)

The budget prioritizes a robust learning system: core technology and infrastructure to ensure reliability and offline capability; human resources to embed coaching and local technical support; and dedicated funds for rigorous monitoring and evaluation. Investments are front-loaded to develop high-quality diagnostics and teacher tools, then taper as capacity transfers to provincial teams. Cost-effectiveness is tracked alongside learning gains to inform scale decisions.

5.2 Budget Breakdown by Category

Budget Category	Amount (R)	% of Total
Technology Development & Infrastructure		
Platform development and customization	7 000 000	15
Hardware and devices	5 800 000	13
Connectivity and data	4 200 000	9
Human Resources		
Project management team	6 000 000	13
Technical support staff	3 300 000	7
Coaching delivery (cluster coaches)	5 000 000	11
Data & Learning Analytics (internal improvement)	2 700 000	6
Capacity Development		
Teacher professional development	3 600 000	8
Learning materials development	2 300 000	5
Operations & Administration		
Project operations	2 100 000	5
Monitoring and evaluation (internal)	1 000 000	2
External evaluation (independent, ring-fenced)	2 750 000	6
TOTAL	45 750 000	100

Note: Totals incorporate a 10% contingency reserve distributed proportionally across categories. The Human Resources line “Data & Learning Analytics (internal improvement)” funds rapid learning and product improvement; formal M&E deliverables (indicators, impact estimation, external reviews) are costed only under the Monitoring and Evaluation headings. See Risk and Appendix for contingency governance.

5.3 Annual Budget Distribution

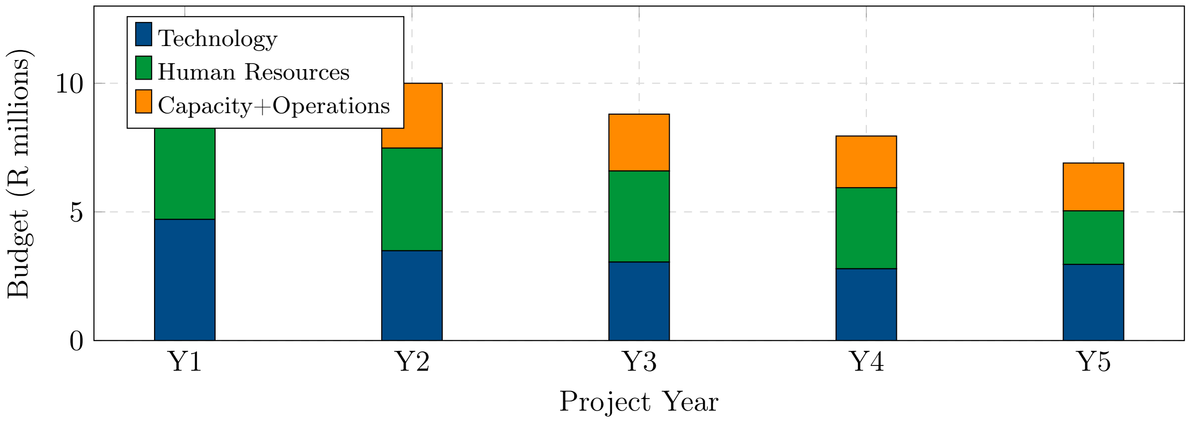


Figure 2: Annual budget distribution aligned to category totals. (Each stack includes its proportional contingency; no separate "Contingency" bar.)

6 Monitoring and Evaluation

6.1 Results Chain Framework

Theory of Change

IF teachers receive real-time diagnostic data and AI-generated pedagogical guidance, **AND** are supported to implement data-driven instruction, **THEN** learners will receive targeted support addressing their specific learning gaps, **LEADING TO** improved mathematics achievement and reduced educational inequality.

The monitoring and evaluation (M&E) strategy links inputs to activities, outputs, outcomes, and impact through a transparent theory of change. Routine platform analytics, teacher practice observations, and standardized assessments provide a triangulated evidence base. Interim learning reviews enable adaptive management while maintaining methodological rigor for summative impact estimation and cost-effectiveness analysis. We will pre-register the protocol and publish a pre-analysis plan prior to baseline measurement on an open registry (e.g., EGAP/OSF), and archive analysis code and redacted data products.

6.2 Key Performance Indicators

Indicator	Type	Baseline	Target (Y5)
Primary KPIs (Pilot-controlled outcomes)			
Learner achievement (IRT scale gain, SD)	Outcome	0.00 SD	+0.25 SD
Teacher practice rubric (TL-DL v1.1)	Outcome	2.1/5	4.2/5
Cost per 0.1 SD improvement (R/learner)	Efficiency	N/A	$\leq$ Y3 benchmark
Secondary KPIs (Equity)			
Achievement gap (Q1 vs Q5)	Outcome	35pp	15pp
Fidelity Metrics			
Learner engagement index	Fidelity	45%	85%
Teacher dashboard weekly active rate	Fidelity	N/A	$\geq$ 80%
Platform availability (Operational SLO)	Ops Metric	N/A	$\geq$ 99.5%
Output Level			
Teachers trained	Output	0	450
Diagnostic assessments completed	Output	0	75,000
Peer-reviewed research papers published (per year)	Output	0	2
Postgraduate students funded (Masters/PhD)	Output	0	5
Conference presentations (cumulative)	Output	0	$\geq$ 10
Contribution (System-level; not primary KPIs)			
Grade 3 numeracy pass rate (province-wide)	Contribution	41%	Informative only

6.3 Evaluation Design

6.3.1 Design

We lock the evaluation to a stepped-wedge cluster randomized trial (CRT) with schools as clusters and three activation waves (1:1:1 allocation). All schools contribute pre- and post-crossover observations; the intention-to-treat (ITT) effect is identified from within- and between-school contrasts across steps.

6.3.2 Matching and Stratification

Prior to randomization, schools will be stratified on variables predictive of outcomes and feasible to measure pre-intervention: school quintile classification (Q1–Q3 vs Q4–Q5; no-fee/fee), urban/rural, language of learning and teaching (Sesotho/Afrikaans/English), enrollment size ($\leq$500 vs >500), connectivity readiness (baseline bandwidth/offline dependence), and baseline mathematics attainment (IRT pretest or most recent Grade 3 pass rate). Randomization to waves will occur within strata using blocked assignment.

6.3.3 Randomization Process

Allocation will be generated by a reproducible script (R) with a fixed random seed, witnessed by an independent methods advisor. The seed, script, and allocation list will be archived with a timestamp in the registry supplement. Allocation will be concealed until announcement to schools; the schedule is fixed and will not be altered based on interim results.

6.3.4 Wave Timing and Measurement

Baseline (T0) is collected for all schools at the end of Year 1. Wave 1 activates in Y2 T1, Wave 2 in Y3 T1, and Wave 3 in Y4 T1. Standardized IRT assessments are administered annually (baseline plus Y2–Y5). Teacher practice (TL-DL v1.1) is observed termly, and fidelity metrics (engagement, WAU, uptime) are logged continuously and summarized termly. All schools contribute outcome data through Y5 for the primary analysis.

6.3.5 Contamination Controls

We gate platform access by school and start date, provision accounts only at wave activation, and separate PLC/coaching by wave. Coaches are assigned to treated or not-yet-treated schools but not both in the same term; materials that could leak content are released after crossover. Usage telemetry flags cross-school account use; any cross-exposure will be logged and reported in the CONSORT extension diagram for stepped-wedge trials.

6.3.6 Analytic Model (Primary)

The primary estimand is the ITT average treatment effect. We fit a linear mixed-effects model for continuous IRT scale scores with fixed effects for step/time and treatment, pre-specified covariates (strata indicators, baseline attainment), and random intercepts for school (and learner where repeated measures permit). Cluster-robust (school-level) standard errors and Satterthwaite degrees of freedom will be used. Teacher-practice and fidelity outcomes use analogous generalized mixed models appropriate to scale. Pre-specified heterogeneity includes quintile, gender, and language. Sensitivity analyses include a cluster-level Hussey–Hughes model and GEE with robust sandwich variance.

6.3.7 Pre-analysis Plan

We will register the design and analysis plan on an open registry (EGAP/OSF) *before baseline data collection*. The PAP will include: outcomes and scales, primary estimand and model, covariates, handling of missing data, multiplicity control, interim reporting rules, randomization script and seed, and a commit-hash of analysis code. The registry entry will be time-stamped; any deviations will be logged via registered updates.

We blend formative learning with summative accountability. Formative components emphasize rapid feedback, continuous stakeholder engagement, and data dashboards to keep implementation responsive. Summative components include annual interim impact analyses aligned to the stepped-wedge rollout and independent external reviews in Years 3 and 5. Cost-effectiveness will be reported as Rand per 0.1 SD improvement with sensitivity analyses (±20% on costs and effects).

7 Risk Mitigation

7.1 Risk Assessment Matrix

Key risks include infrastructure limitations, teacher resistance to change, funding variability, technology failures, and policy shifts. Our strategy prioritizes offline-capable tools, phased adoption with peer champions, diversified and staged financing, technical redundancy with local support capacity, and proactive multi-stakeholder engagement. Risks and mitigations are revisited at each governance checkpoint and adjusted in response to evidence from implementation.

Risk	Probability	Impact	Mitigation Strategy
Infrastructure limitations	High	High	Offline-capable solutions, phased rollout
Teacher resistance	Medium	High	Incentives, peer champions, gradual adoption
Funding shortfalls	Medium	High	Diversified funding, cost optimization
Technology failures	Low	Medium	Redundancy, local support capacity
Policy changes	Low	Medium	Multi-stakeholder engagement

Each risk has an assigned owner, early-warning indicators, and decision thresholds for mitigation activation; the register is reviewed at each governance checkpoint.

7.2 Contingency Planning

Contingency Reserve: 10% of total budget (R 4,575,000) allocated for unforeseen challenges and opportunities

8 Ethics and Data Protection

8.1 Ethical Approval and Oversight

Governance

Ethical clearance will be obtained from the University of the Free State Research Ethics Committee and the Free State Department of Education. An independent advisory panel will review evaluation protocols, consent processes, and safeguarding practices annually.

8.2 Informed Consent and Safeguarding

We will secure parental or guardian consent and learner assent for all research-related data collection, and obtain teacher and school consent for classroom observations and usage analytics. Safeguarding protocols aligned with DBE policies will guide conduct in schools, with clear incident reporting and escalation procedures to protect minors and staff.

8.3 POPIA Compliance and Security

Data processing follows POPIA's principles of purpose limitation and data minimization. Reporting uses pseudonymized and aggregated datasets, with no learner-identifiable data in analytics outputs. All personal data will be processed and stored in a South Africa cloud region with key management via KMS/HSM. Access is role-based, least-privilege, and logged for audit. Cross-border transfers will not occur without POPIA-compliant safeguards and Data Processing Agreements (DPAs). Regular security reviews and penetration tests will be conducted.

8.4 Data Management Plan

Data will be hosted securely with defined retention schedules, and de-identified prior to research use. The research team will operate with tiered access permissions and maintain data dictionaries and version control for instruments and code. De-identified, aggregated findings and reusable code will be shared openly where appropriate to advance the field and support transparency.

8.5 Algorithmic Fairness and Model Governance

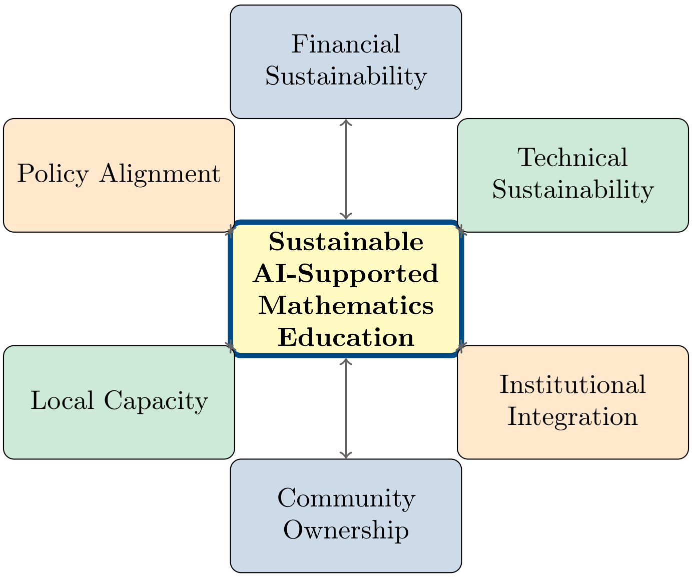
Model development and deployment will follow documented *model cards* describing data sources, training procedures, intended use, and limitations. We will conduct pre-deployment bias audits by subgroup (quintile, language, gender) and monitor post-deployment drift and fairness metrics. An escalation protocol will pause or roll back models that exhibit harmful bias or performance degradation, with remediation tracked through change logs and governance reviews. We will also use performance banding to ensure subgroup comparisons at similar proficiency levels and to avoid aggregation artifacts in class-imbalanced settings.

8.6 Data Protection Impact Assessment

A POPIA-aligned Data Protection Impact Assessment (DPIA) will be completed before deployment and reviewed annually.

9 Sustainability and Scale

9.1 Sustainability Framework



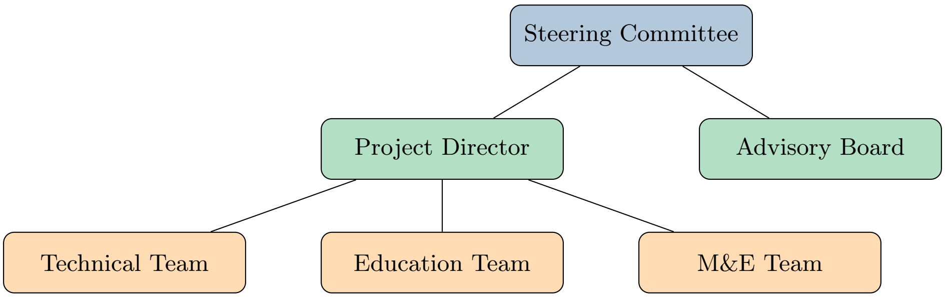
9.2 Scale-Up Pathway

Provincial to National Scale

1. **Years 1-2**: Proof of concept in 3-5 Free State schools
2. **Years 3-4**: Provincial adoption across 15 schools
3. **Year 5**: Provincial policy integration and preparation for expansion
4. **Post-project**: National rollout through DBE partnership

10 Project Team and Governance

10.1 Governance Structure



Governance cadence: The Steering Committee meets quarterly; the Advisory Board meets biannually. Public summary minutes will be published to promote transparency and accountability.

10.2 Key Personnel

Position	Key Responsibilities	FTE
Project Director	Overall project leadership and stakeholder management	1.0
Technical Lead	Platform development and maintenance	1.0
Education Specialist	Curriculum alignment and teacher support	1.0
M&E Coordinator	Data collection and impact assessment	0.8
Training Coordinators	Teacher professional development (x3)	3.0
Research Associates	Data analysis and reporting (x2)	2.0

11 Expected Outcomes and Impact

11.1 Short-term Outcomes (Years 1-2)

✓ Functional AI platform deployed

✓ 100+ teachers trained

✓ Baseline data established

✓ Proof of concept validated

✓ Stakeholder buy-in secured

✓ Initial learning improvements

11.2 Medium-term Outcomes (Years 3-4)

- 20% improvement in numeracy scores
- 1,500 learners actively engaged
- 15 schools fully integrated
- Best practices documented
- Partnership model established

11.3 Long-term Impact (Year 5 and beyond)

Transformative Impact

1. **Educational Transformation**: Fundamental shift to data-driven, personalized mathematics instruction
2. **Equity Enhancement**: Significant reduction in achievement gaps across socioeconomic groups
3. **System Strengthening**: Enhanced teacher capacity and institutional effectiveness
4. **Economic Returns**: Improved STEM pipeline contributing to economic development
5. **Knowledge Generation**: Evidence base for AI in education policy development

12 Conclusion

Investment Opportunity

This proposal presents a carefully designed, evidence-based intervention that addresses critical mathematics education challenges in the Free State province. Through strategic deployment of AI-supported pedagogy, comprehensive teacher development, and rigorous evaluation, the project promises to deliver:

- **Measurable Learning Gains**: Targeted +0.25 standard-deviation improvement on the IRT scale
- **Enhanced Equity**: Reduced achievement gaps by 50%
- **Innovation Leadership**: Position South Africa as leader in EdTech for development
- **Sustainable Change**: Self-sustaining model for national scale
- **Value for Money**: Cost transparency at R 8 632/learner-year (∼R 22 875/unique learner)

We invite ETDP-SETA to partner with us in this transformative initiative that aligns with national priorities, leverages cutting-edge technology, and promises lasting impact on South Africa's education landscape.

A Power Analysis

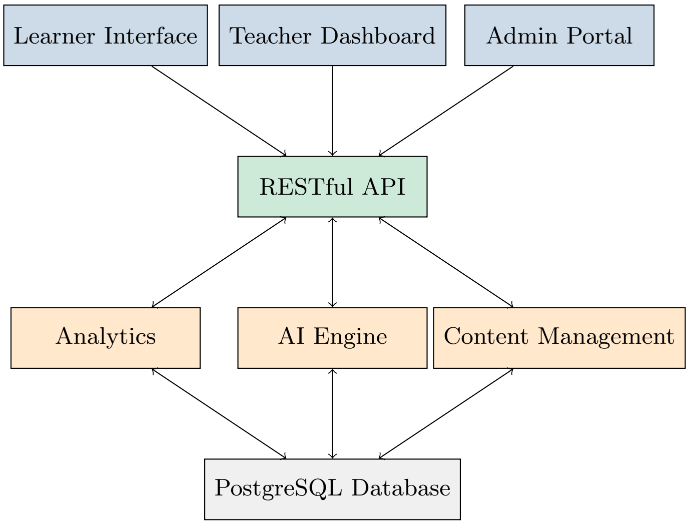
We assume a school-level intraclass correlation (ICC) of 0.12 for mathematics, an average cluster size of 120 learners, and baseline–endline score correlation of 0.60. With 15 clusters in a stepped-wedge design rolled out in three waves, simulation-based power analyses (10,000 replications) indicate approximately 80% power to detect a 0.22–0.25 SD effect under conservative attrition scenarios. If ICC ≥ 0.18 or attrition $\geq 20\%$, we will either recruit up to five additional schools and randomize them to a later wave or extend the wedge by one step to maintain target power while preserving the CRT design. Power simulations assume three steps (Y2 T1, Y3 T1, Y4 T1) with annual measurements at T0–T5, equal cluster sizes within strata, and the canonical Hussey–Hughes specification with a linear time effect tied to the stepped-wedge schedule. We assume a coefficient of variation of cluster sizes (CV) of 0.20 and near-zero rollover contamination between waves, given access scoping and phased enablement controls.

B Measurement and Psychometrics

We will maintain a calibrated item bank aligned to CAPS with annual anchor sets for equating, monitor parameter drift, and conduct DIF testing across gender, language, and quintile. Scale scores will be reported with cluster-robust confidence intervals. Teacher practice measures (e.g., use of diagnostics, feedback quality) will use pre-specified rubrics with double-coding and inter-rater reliability targets of $\kappa \geq 0.7$.
Power simulations implemented in R using `swCRTdesign` and `simr`; scripts will be stored in the study repository for audit and reuse.

C Technical Specifications

C.1 Platform Architecture



C.2 Offline-First Architecture and Edge Inference

The platform is an offline-first PWA that uses Service Workers for caching, background sync for deferred uploads, and conflict resolution policies on reconnect. Core inference models are quantized (int8) for on-device execution with TensorFlow Lite or ONNX Runtime Mobile when available, falling back to server inference or rule-based recommendations if required. Client queues are idempotent; server endpoints enforce idempotency to avoid duplication on retries.

C.3 Data Model and MLOps

Event-level telemetry captures assessment responses, hint requests, and time-on-task with a privacy-preserving schema using pseudonym keys and scoped access by school/teacher/learner. Models and datasets are versioned; drift detection and shadow deployments precede canary releases in a small set of schools. A model registry and reproducible training pipelines ensure auditability.

C.4 SLOs and Observability

We target a core service SLO of $\geq$ **99.5%** availability, contextualized by offline capability for learners and teachers. Health checks, tracing, and alerting (OpenTelemetry, Prometheus, Grafana) support incident response and error budgets aligned with the SLO.

C.5 Cloud Region and Data Residency

All personal data are stored and processed in a South Africa cloud region. No cross-border transfers occur without POPIA-compliant safeguards, DPAs, and documented risk assessments.

C.6 Device Management and Resilience

Devices will be managed via MDM with kiosk-mode profiles, remote wipe, and update policies. A spares pool ($\geq$ 5%) covers breakage and loss. Lost/stolen procedures and inventory audits are documented with district leads.

C.7 Model Specifications and Targets

IRT: 2PL for multiple-choice items; 3PL for selected items to account for guessing. Calibration in `mirt`/`TAM` with annual anchor-based equating; local dependence checks (Yen's Q3 ≤ 0.20).
Polytomous: Graded Response Model (GRM) for rubric-scored items.
Knowledge tracing: Compact DKT/GRU or AKT models with targets of AUC ≥ 0.70 and ECE ≤ 0.05.
Edge budgets: Per-model $\leq$ 8 MB; cumulative pack $\leq$ 25 MB per device/language; p95 on-device inference $\leq$ 50 ms.
Fairness: Pre/post-deploy DIF (Mantel–Haenszel and IRT-LR), subgroup error-rate parity by quintile/language/gender; mitigations via item revision, reweighting, or pathway caps documented in model cards.

C.8 Offline Sync and Reconciliation

Assessment telemetry is event-sourced and idempotent. Clients queue NDJSON payloads (`gzip`, 512 KB chunks, max batch 2 MB) with idempotency keys; the server merges streams (no overwrite for assessment events), and uses vector clocks with last-writer-wins for profile fields. Multi-week offline operation is supported; read models are eventually consistent and scoped at school level, refreshed on sync. Expected volume is 0.25–0.5 MB/learner/day (compressed); at 2,000 learners this is ≈0.5–1.0 GB/day system-wide (35–70 MB/day per school across 15 schools). Sync completion SLO: p95 $\leq$ 30 s on school Wi-Fi, $\leq$ 2 min on 3G.

C.9 Resilience and Security SLOs

RPO 24 h and RTO 4 h with cross-AZ replicas and monthly restore drills. Identity provider via Cognito or Keycloak with OAuth2+PKCE; API Gateway rate-limiting (token-bucket) per user and per school; AWS WAF and Shield at the edge. Encryption at rest (AES-256 with KMS CMKs) and TLS 1.2+ in transit; field-level encryption for national ID numbers; secrets in AWS Secrets Manager.

C.10 MLOps and Monitoring

Feast feature store (offline Parquet, online Redis); MLflow registry with promotion gates (performance and fairness). Drift monitors (PSI, KL) trigger alerts when PSI > 0.2 for three consecutive days. Champion/challenger with 10% shadow canaries precedes full promotion; nightly retrains on a T4/A10-class GPU, budgeted $\leq 2\,\text{h}$/job.

C.11 Performance Targets

Throughput 200 RPS steady (burst 500 RPS for 1 min); error budget 0.5%. API latency targets: p50 80 ms / p95 250 ms / p99 400 ms. CDN TTFB p95 $\leq$ 150 ms. Sync completion p95 as specified in Offline Sync and Reconciliation.

C.12 Integrations and SA-SAMS

Phase 1 provides CSV import (learners, classes, teachers) with a documented schema map. Phase 2 adds REST connectors and a published OpenAPI for roster, content, and telemetry services. Static assets are served via CDN. Communication fallbacks include SMS for nudges and USSD for lightweight attendance/alerts; adaptive diagnostics do not use USSD.

C.13 Skills Transfer and Technical Debt

We will pair provincial staff with engineers for co-delivery, provide admin training and runbooks, and reserve a technical-debt budget ($\geq$ 8% of development effort) for refactors and maintenance to ensure sustainable handover.

C.14 Technology Stack

Component	**Technology**
Frontend	React.js, Progressive Web App (Service Workers, background sync)
Backend	Node.js, Express.js
AI/ML	Python, TensorFlow/PyTorch; TensorFlow Lite/ONNX Runtime Mobile for edge; mlflow
Database	PostgreSQL (row-level security), Redis
Infrastructure	AWS (Africa – Cape Town), Docker, Kubernetes, Terraform, ArgoCD (GitOps)
Analytics	PostgreSQL + DuckDB for OLAP; Metabase/Superset for dashboards
Observability	OpenTelemetry, Prometheus, Grafana, Loki
Security	OAuth 2.0 (+ PKCE), SSL/TLS, POPIA compliance, KMS-managed keys

C.15 IP and Licensing

To maximize reuse and sustainability, platform source code will be released under `Apache-2.0`, and teacher-created materials and training resources under `CC BY-SA 4.0`, subject to final Data Sharing and IP agreements with DBE and ETDP-SETA. Model cards, item parameter documentation, and process artifacts will be openly published where compatible with assessment security.

D Detailed Budget Justification

D.1 Technology Development Costs

Item	Unit Cost	Quantity	Total (R)
Platform Development			
Requirements analysis	150,000	1	150,000
UI/UX design	250,000	1	250,000
Frontend development	1,200,000	1	1,200,000
Backend development	1,500,000	1	1,500,000
AI module development	2,000,000	1	2,000,000
Testing and QA	400,000	1	400,000
Infrastructure			
Cloud hosting (annual)	240,000	5	1,200,000
Security and compliance	300,000	5	1,500,000
Backup and disaster recovery	60,000	5	300,000
Subtotal			**8,500,000**

E Readiness Attachments

The following attachments demonstrate implementation readiness. Redactions are applied where appropriate to protect privacy and procurement-sensitive information.

- School readiness and timetabling rubric (template) with three anonymised completed examples.
- Coaching delivery pack: term-level coach schedules, coaching tracker template, and cluster workshop agendas.
- Measurement pack: TL–DL v1.1 rubric and moderation protocol; IRT assessment blueprint with example item shells; registry/PAP skeleton with fields pre-specified.
- Data protection and security: POPIA DPIA template, role-based access matrix, DPA template, SLO/uptime runbook, and incident response playbook.
- Technology operations: platform architecture, offline sync and reconciliation runbook, MDM configuration profiles, and device inventory log template.
- Governance and risk: Steering Committee ToR, RACI and escalation path, risk register with thresholds and early-warning indicators.

F References and Bibliography

Key References

- Department of Basic Education. (2023). *National Assessment Results 2023.* Pretoria: DBE.
- ETDP-SETA. (2024). *Strategic Plan 2020-2025.* Johannesburg: ETDP-SETA.
- IEA. (2023). *TIMSS 2023 International Results in Mathematics.* Boston: TIMSS & PIRLS International Study Center.
- Mosia, M. (2025). *Theory of Change and Logic Model: AI-Supported Evidence-Based Mathematics Pedagogy.* University of the Free State.
- National Planning Commission. (2012). *National Development Plan 2030.* Pretoria: The Presidency.
- OECD. (2023). *PISA 2022 Results: Mathematics Performance.* Paris: OECD Publishing.
- Reddy, V., et al. (2023). *TIMSS 2023: South African Learners' Achievement.* Cape Town: HSRC Press.
- Statistics South Africa. (2024). *General Household Survey 2023.* Pretoria: Stats SA.
- UNESCO. (2023). *Global Education Monitoring Report 2023.* Paris: UNESCO.
- World Bank. (2024). *EdTech in Developing Countries: Evidence Review.* Washington: World Bank Group.
- Department of Basic Education & Free State DoE. (2019). *Free State Grade 3 Systemic Evaluation Report.* Unpublished provincial data.
- ETDP-SETA. (2024). *Sector Skills Plan 2025–2030.* Johannesburg: ETDP-SETA.
- ETDP-SETA. (2025). *Landscape Review of Mathematics Education in South Africa: Summary Report.* Johannesburg: ETDP-SETA.
- Pane, J. F., et al. (2014). *Effectiveness of Cognitive Tutor Algebra I at Scale.* RAND Corporation.
- OECD. (2020). *AI in Education: Policy Challenges and Recommendations.* Paris: OECD Publishing.
- Woolf, B. P., et al. (2009). *Building Intelligent Tutoring Systems.* San Francisco: Morgan Kaufmann.
- Protection of Personal Information Act (POPIA). (2013). Republic of South Africa.
- Academy of Science of South Africa (ASSAf). (2018). *Ethics in Education Research: Guidelines.* Pretoria: ASSAf.
- ASSAf. (2023). *POPIA Code of Conduct for Research.* Pretoria: ASSAf.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918.
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.

Glossary

- CAPS: Curriculum and Assessment Policy Statement (South Africa)
- ICC: Intraclass Correlation Coefficient
- ITT: Intention-to-Treat (analysis)
- DIF: Differential Item Functioning
- RCT: Randomized Controlled Trial (cluster, stepped-wedge variant)
- SLO: Service Level Objective

Contact Information

Prof. Moeketsi Mosia
Mathematics ETDP-SETA Research Chair
University of the Free State
+27 51 401 3000
etdp.research@ufs.ac.za
www.ufs.ac.za/etdp-mathematics

Submission Date: September 19, 2025

Transforming Mathematics Education

One Learner, One Teacher, One School at a Time