



EVIDENCE AND EVALUATION | 2026 PRACTITIONER EDITION

STEM Programme Evaluation and Impact Framework

A disciplined route from programme theory and indicator design to credible claims, adaptive decisions and responsible reporting.

Jonathan Kennedy Sowah

AI Engineer and K-12 STEM Systems Builder
Ghana | Africa | jonathansowah.com

For programme leaders, monitoring and evaluation teams, funders, researchers, schools and public institutions.

FIELD-TESTED THINKING | PRACTITIONER TEMPLATES | EVIDENCE ANCHORS

What this manual helps you decide

Move from idea to explicit decisions, owners, evidence and next actions.

1	Make the causal logic visible	Evaluation begins by explaining how activities are expected to produce capability and under which conditions.
2	Match claims to the level of evidence	Reach, experience, capability, transfer, progression and durability answer different questions.
3	Define indicators before collecting data	A number becomes useful only when its meaning, source, cadence, quality and decision purpose are stable.
4	Use proportionate methods and protect participants	Evaluation design should be strong enough for the decision while respecting learners, schools and communities.
5	Turn evidence into decisions, not decoration	Evaluation creates value when findings alter design, funding, support, scale or stop decisions.
6	90-day application	Use 90 days to clarify programme theory, repair indicators, test instruments and establish a learning cadence.

Use the manual as a decision cycle

1 OBSERVE	2 DECIDE	3 ASSIGN	4 REVIEW
Start with the ordinary operating condition, not the launch-day exception.	Record the choice, the alternatives rejected and the assumption carrying the most risk.	Name one accountable owner, the supporting actors and the evidence they must produce.	Set a date and define what evidence will trigger continuation, repair, redesign or stop.

A completed template is not evidence of success. Pair every decision with an observable artefact, a named owner and a scheduled review.

Make the causal logic visible

Evaluation begins by explaining how activities are expected to produce capability and under which conditions.

A theory of change should expose assumptions, mechanisms, alternative explanations and failure points - not decorate a proposal.

DESIGN MOVE	WHAT IT REQUIRES	EVIDENCE TO KEEP
Problem	Define whose problem, compared with what baseline, in which context and why it persists.	Problem analysis and disaggregated baseline
Mechanism	Explain how dosage, pedagogy, tools, relationships and support change participant behaviour.	Causal pathway with testable assumptions
Outcome	Specify observable changes in learner, teacher and institutional capability.	Outcome definitions and time horizon
Context	Record infrastructure, policy, staffing, selection and implementation conditions that affect results.	Context register and comparison notes

Decision questions

1. What must be true for the activity to cause the intended outcome?
2. Which plausible alternative explanation could produce the same observation?
3. Where is the pathway most likely to break?
4. What early signal would justify adaptation?

Decision gate

PROCEED WHEN	PAUSE OR REDESIGN WHEN
The causal chain, assumptions and decision uses are explicit before indicators are selected.	The theory jumps directly from activities to transformation or treats context as background detail.

Field application

Take the programme's headline outcome and draw every required step backwards to an activity. Mark each unsupported assumption and name evidence needed.

Match claims to the level of evidence

Reach, experience, capability, transfer, progression and durability answer different questions.

A credible report climbs only as high as its evidence. It does not convert attendance into learning or short-term performance into lasting system change.

DESIGN MOVE	WHAT IT REQUIRES	EVIDENCE TO KEEP
Reach	Establish unique participants, denominator, dosage, attendance threshold and duplication method.	Clean register and exposure calculation
Experience	Examine feasibility, quality, safety, relevance and participant response.	Observation, feedback and incident evidence
Capability	Use authentic performance to determine what participants can do and explain.	Task, rubric and moderated judgement
Transfer and durability	Test new-context use and what remains when support, novelty or incentives reduce.	Follow-up performance and operating evidence

Decision questions

1. Which rung does each public claim occupy?
2. Does the evidence directly test that rung?
3. What denominator and limitation accompany the number?
4. What evidence would contradict the preferred story?

Decision gate

PROCEED WHEN	PAUSE OR REDESIGN WHEN
Claim wording, method, source and limitation are linked in a claim register.	Reports use reached, trained, empowered and transformed as interchangeable terms.

Field application

Underline every outcome verb in the latest report. Classify its evidence level and rewrite any sentence that reaches above its supporting data.

Define indicators before collecting data

A number becomes useful only when its meaning, source, cadence, quality and decision purpose are stable.

Good indicators reduce ambiguity and trigger action. They are not selected merely because a dashboard can display them.

DESIGN MOVE	WHAT IT REQUIRES	EVIDENCE TO KEEP
Definition	Specify numerator, denominator, unit, threshold, disaggregation and time period.	Approved indicator reference sheet
Source	Name the instrument, collection owner, system location and verification method.	Data lineage and access record
Quality	Test completeness, consistency, validity, duplication, missingness and assessor reliability.	Quality check and correction log
Decision	State what action the indicator informs and when it is reviewed.	Decision calendar and accountable owner

Decision questions

1. Can two analysts calculate the same value independently?
2. What behaviour might the indicator unintentionally encourage?
3. Which groups disappear inside the average?
4. What decision changes if the value crosses the threshold?

Decision gate

PROCEED WHEN	PAUSE OR REDESIGN WHEN
Indicators have reference sheets, owners, quality tests and decision links.	Definitions change between reports or metrics are collected without a plausible action.

Field application

Give an indicator name to three colleagues and ask each to calculate it from raw records. Any disagreement reveals a definition or data-lineage defect.

Use proportionate methods and protect participants

Evaluation design should be strong enough for the decision while respecting learners, schools and communities.

Method choice depends on the question, stakes, feasibility and ethical burden. More data is not automatically better evidence.

DESIGN MOVE	WHAT IT REQUIRES	EVIDENCE TO KEEP
Design	Choose baseline, repeated measures, comparison, contribution analysis or qualitative inquiry deliberately.	Evaluation protocol and design rationale
Sampling	Define population, selection, attrition, subgroup power and limits to generalisation.	Sampling frame and response analysis
Triangulation	Combine performance, observation, administrative data and participant explanation where useful.	Convergence and contradiction matrix
Protection	Minimise data, secure consent or assent, limit access and plan safe reporting.	Ethics, safeguarding and data plan

Decision questions

1. What decision requires this evidence?
2. What is the least burdensome credible design?
3. Who may be harmed, excluded or misrepresented?
4. How will negative or unexpected findings be handled?

Decision gate

PROCEED WHEN	PAUSE OR REDESIGN WHEN
The method can answer the question and the burden is justified and protected.	Evaluation collects sensitive data without necessity or claims causality from weak design.

Field application

For each planned field, write the decision it supports, who can access it and when it will be deleted. Remove fields without a defensible purpose.

Turn evidence into decisions, not decoration

Evaluation creates value when findings alter design, funding, support, scale or stop decisions.

Report uncertainty, counterevidence and context alongside achievement. Publish the decision path, not only attractive results.

DESIGN MOVE	WHAT IT REQUIRES	EVIDENCE TO KEEP
Sensemaking	Review findings with implementers and participants while protecting independent judgement.	Interpretation record and contested findings
Adaptation	Translate evidence into owners, deadlines, conditions and follow-up measures.	Management response and action tracker
Reporting	Separate verified result, estimate, inference, participant view and aspiration.	Claim-evidence register and limitations
Scale	Test quality, equity, cost and operating durability before extrapolating.	Scale decision pack and replication conditions

Decision questions

1. What changed because of the evidence?
2. Which finding is least comfortable but most decision-relevant?
3. Can a reader trace a claim to source, method and limitation?
4. What must be learned before the next funding gate?

Decision gate

PROCEED WHEN	PAUSE OR REDESIGN WHEN
Findings are converted into recorded decisions with review dates.	Evidence is produced after decisions, used only for promotion or stripped of limitations.

Field application

Run a decision meeting where the team must choose continue, repair, redesign or stop. Require one supporting and one opposing piece of evidence for the choice.

Build a decision-ready evidence system

Use 90 days to clarify programme theory, repair indicators, test instruments and establish a learning cadence.

PERIOD	GATE	FIELD WORK	OUTPUT
Days 1-15	Diagnose	Observe ordinary conditions, establish the baseline and name the highest-risk assumptions.	Approved baseline and decision register
Days 16-30	Design	Select a bounded use case, define responsibilities and agree the evidence standard.	Signed design brief and risk controls
Days 31-60	Pilot	Run repeated practice under normal constraints; log failures, adaptations and participant experience.	Pilot evidence pack and repair log
Days 61-75	Review	Compare intended capability with observed performance and test equity, cost and continuity.	Review decision with conditions
Days 76-90	Institutionalise	Assign recurring ownership, budget, cadence and escalation routes before expansion.	Operating plan and next review date

Decision record

Priority decision	_____
Outcome claim	_____
Evidence level	_____
Indicator definition	_____
Data source	_____
Quality risk	_____
Responsible owner	_____
Review date	_____

Do not collect data because it is available. Collect the smallest credible evidence set that protects participants and improves a consequential decision.

Evidence anchors and further reading

The frameworks translate public guidance and implementation evidence into practical decision tools.

SOURCE	RESOURCE	ADDRESS
Ghana Ministry of Education	Education Strategic Plan 2018-2030 (2018)	https://moe.gov.gh/edge/content/uploads/2019/05/Education-Strategic%20Plan-2018-2030.pdf
UNESCO GEM Report	Technology in education: A tool on whose terms? (2023)	https://gem-report-2023.unesco.org/
UNICEF Innocenti	Digital Learning, Real Classrooms: Implementation Research in Ghana (2025)	https://www.unicef.org/innocenti/reports/digital-learning-real-classrooms
African Union	Continental Artificial Intelligence Strategy (2024)	https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy
African Union	Continental Education Strategy for Africa 2026-2035 (2025)	https://static.au.int/en/documents/20250616/continental-education-strategy-africa-2026-2035

Use these sources as starting points. Confirm current law, policy, curriculum and regulator guidance before making consequential institutional decisions. This manual is an implementation aid, not legal, financial or safeguarding advice.

Evidence quality protocol

CHECK	QUESTION	RECORD
Currency	Is the source still current for the decision date, jurisdiction and learner group?	Version, publication date and date checked
Authority	Is it primary guidance, independent research, implementation evidence or vendor material?	Source type and responsible institution
Fit	Do its assumptions match the school, infrastructure, language, staffing and safeguarding context?	Transfer limits and local verification needed
Use	Which decision does this source inform, and what would cause that decision to change?	Decision link, owner and review trigger

Measure capability, transfer and durability

A practical contribution to stronger AI, coding, robotics and STEM education systems in Ghana and Africa.

The value of a framework is not the elegance of its diagram. It is the quality of the decisions, routines and learning that it makes possible.

About Jonathan Kennedy Sowah

Jonathan is a Ghanaian AI engineer and K-12 STEM systems builder working across robotics, coding, applied artificial intelligence, teacher development and context-aware education technology. His wider portfolio has reached 25,000+ learners, trained 2,500+ educators and engaged 100+ schools. KodeVR has more than 10,000 active learners.

He works with schools, governments, companies and development partners to turn high-visibility technology ambitions into practical learning systems that can operate under real African conditions.

Use this work responsibly

Adapt the tools to the programme, participants, institution and regulatory context. Record assumptions. Seek qualified technical, legal, safeguarding, financial and evaluation review where the decision requires it.

Before commissioning implementation support

QUESTION	A STRONG ANSWER SHOULD NAME
What capability must remain after the engagement?	The learner, teacher and institutional behaviours that should continue without external supervision.
What conditions shape delivery?	The ordinary timetable, infrastructure, language, safeguarding, procurement and maintenance realities.
How will quality be governed?	Decision owners, evidence definitions, review cadence, repair route and stop-or-scale thresholds.
What will the institution inherit?	Reusable learning assets, local capability, documented routines, data rights and an affordable continuity plan.

jonathansowah.com | info@jonathansowah.com