Balochistan Winter School Programme

Field Intelligence Report

Combining Field Observations with Classroom Data Analysis

Programme: FLN - Foundational Literacy & Numeracy

Report Date: January 23, 2026

Field Observations: Sabeena Abbasi

Data Analysis: 522 AI + 54 human coach observations + 191 lesson plans

Table of Contents

Executive Summary

This report combines first-hand field observations from teacher and coach interviews with quantitative analysis of 576 classroom observations. The findings reveal a programme that is working well within its current constraints, but faces significant questions about sustainability and scalability.

The Bottom Line

Finding	Implication
Teachers respond positively to AI observations	Genuine appreciation OR incentive-driven compliance
PKR 35,000 incentive drives high compliance	Teachers complete all uploads and lesson plans
Human coaches are essential	Troubleshooting, relationship-building, always available
One-week training was effective	Both teachers and coaches credit it for success
Phonics understanding remains weak	Teachers use letters instead of sounds—need continuous reinforcement
Implementation gap persists	87-91% gap between lesson plans and actual practice
Kids enjoy the programme	Positive engagement during winter school
Scalability to main season is uncertain	Curriculum pressure, large classes, no incentive

Key Metrics at a Glance

Metric	Value	Source
AI Observations	522	Database
Human Coach Observations	54	Database
Teachers with Both	34	Cross-reference
Teacher Talk Time	81.8%	AI analysis

Student Talk Time	5.7%	AI analysis
Open-Ended Questions	13%	AI analysis
Student-Generated Questions	0%	AI analysis
Implementation Gap	87-91%	Plans vs observations
Teacher Incentive	PKR 35,000	Programme design

Part 1: Programme Context

The Winter School Model

Element	Detail
Programme	FLN (Foundational Literacy & Numeracy) Winter School
Location	Balochistan province, Pakistan
Duration	Winter vacation period
Incentive	PKR 35,000 (~$125 USD) per teacher
Requirements	Complete all lesson plans + upload all observations
Training	One-week intensive onboarding

Why Winter School Works

Teachers explicitly stated that FLN programmes work better during vacations because:

- No curriculum completion pressure — Teachers have protected time
- Smaller class sizes — Winter school has fewer students
- Focused attention — Only FLN content, no competing subjects
- Incentive structure — Clear payment tied to completion

> *"During the regular session, we are mostly worried about completing the curriculum. The attendance —there are too many students in the classroom for us to focus on foundational skills." — Teacher Feedback*

The Vibe: Experienced and Motivated Teachers

Field Observation: *"There was a different vibe. The teachers were well-experienced and highly motivated."*

However, there is an important caveat: teachers were explicitly told that they must complete all lesson plans and upload all observations to receive their PKR 35,000 payment. This incentive likely explains much of the observed motivation.

Part 2: The Incentive Question

What the Incentive Structure Looks Like

Requirement	Tied to Payment?
Complete all lesson plans	✅ Yes
Upload all observations	✅ Yes
Attend training	✅ Yes
Teaching quality scores	❌ No

Result: 100% compliance with administrative requirements, regardless of teaching quality.

Teacher Attitudes: Genuine or Strategic?

What Teachers Said:

- "We like the AI observations"
- "We agree with the feedback"
- No negative comments about the system
- Initially skeptical, now positive

The Interviewer's Assessment:

> *"My hunch is more to do with the incentive. However, even after probing, I didn't see any teacher saying anything negative about the observation."*

Possible Interpretations:

- Teachers genuinely find it useful
- Teachers don't want to jeopardize payment
- Social desirability bias
- Fear of consequences for criticism

The Indirect Benefits (Coach Perspective)

Coaches reported that even if compliance is incentive-driven, the behavioural effects are real:

Indirect Benefit	Explanation
Teachers more polite to students	Knowing they're being recorded
Closer adherence to lesson plans	Following instructions more carefully
More student participation	Creating opportunities for engagement
Use of recommended resources	Pebbles, manipulatives, materials

> *"We should enforce it during the main season as well." — Coach Recommendation*

Key Insight: Even compliance-driven observation creates positive behaviour change. The mechanism (incentive vs intrinsic motivation) may matter less than the outcome (improved practice).

Part 3: The Role of Human Coaches

First Two Weeks: Technical Support

During the programme's initial phase, coaches primarily focused on:

- App troubleshooting
- Platform navigation help
- Upload problem-solving
- Technical support

This was essential. Without coach support, many teachers would have struggled with the technology.

What Teachers Value About Coaches

Aspect	Teacher Feedback
Availability	"Coaches are always there"
Responsiveness	"They respond promptly"
Communication	Teachers reach out via WhatsApp
Relationship	Teachers feel supported

> *"Teachers would reach out to coaches and call them on WhatsApp, showing their interest." — Field Observation*

The Human-AI Division of Labor

Human Coaches Do	AI Observation Does
App troubleshooting	Consistent feedback on every lesson
Relationship building	Quantitative metrics

Contextual judgment	Evidence-based assessment
Emotional support	Scalable coverage (9.7× more observations)
Complex problem-solving	Objective, unbiased evaluation

Observation Frequency Comparison

Type	Count	Frequency
AI Observations	522	Can be every lesson
Human Observations	54	Limited by coach capacity
Ratio	9.7×	AI is 9.7× more frequent

Human coaches cannot observe every lesson. AI fills the coverage gap while humans provide depth and relationship.

Part 4: Training Effectiveness

One-Week Training: A Critical Success Factor

Both teachers AND coaches credited the one-week training for programme success.

What Training Accomplished:

Component	Impact
Programme explanation	Teachers understood FLN goals
App onboarding	Teachers could use the platform
Rules clarification	Clear expectations set
FLN concepts	Drilling, blending, segmenting explained
Relationship building	Coach-teacher bond established

The Phonics Gap: Training vs. Practice

Critical Field Observation: Despite training, teachers still don't fully understand phonics.

What Training Taught	What Teachers Actually Do
Use sounds (phonics)	Use letter names
Focus on blending sounds	Focus on spelling letters
"sss" for S	"ess" for S
Sound-based reading	Alphabet-based reading

Example: Instead of teaching the sound "sss" for the letter S, teachers teach "ess" (the letter name).

Why This Matters

> *"One week of training can get them started, but continuous reinforcers are important. Once teachers get the hang of it, only then can they pass on the knowledge." — Field Insight*

Teacher Acknowledgment:

- "We didn't know what phonics was before training"
- "We now understand its importance"
- "We see how it affects reading development"

But: Understanding importance ≠ implementing correctly. The skill gap remains.

Part 5: Implementation Gap Analysis

The Core Finding

> **Key Insight:** The problem is not knowledge—it's implementation. 100% of lesson plans contain open-ended questions. Only 13% of classrooms use them. 97% of lesson plans prescribe collaboration. Only 6% of classrooms implement it.

Plans vs. Reality

Practice	In Lesson Plans	Observed	Gap
Open-Ended Questions	100%	13%	-87%
Student Collaboration	97%	6%	-91%
Real-Life Connections	100%	16%	-84%
Student Question Prompts	16%	0%	-16%

The Lesson Plans Are Good

Analysis of 191 lesson plans shows they include:

- ✅ Clear learning objectives (100%)
- ✅ Open-ended questions (100%)
- ✅ Collaboration instructions (97%)
- ✅ Real-life connections (100%)
- ✅ Concrete manipulatives (pebbles, sticks, beans)
- ✅ Differentiation by student level
- ✅ Concrete → Pictorial → Abstract progression

What Teachers Actually Do

Strong Areas (Teacher-Centered):

- Classroom management: 90%
- Stating learning objectives: 84%
- Following content sequence: 82%
- Maintaining safe environment: 90%
- Asking questions: 85% (but mostly closed-ended)

Weak Areas (Student-Centered):

- Implementing pair/group work: 6%
- Asking open-ended questions: 13%
- Checking for understanding: 44%
- Making real-life connections: 16%
- Allowing student choice: 44%
- Lesson closure/summary: 41%

Positive Sign: Resource Usage

Field Observation: Teachers ARE using the resources in Maths lesson plans:

- Pebbles for counting ✅
- Sticks for manipulatives ✅
- Other hands-on materials ✅

Implication: Teachers follow the concrete parts of lesson plans (materials) but struggle with the pedagogical parts (questioning, collaboration).

Part 6: The Classroom Reality

Talk Time Distribution

Speaker	% of Class Time	Minutes (40-min lesson)
Teacher	81.8%	33 minutes
Students	5.7%	2 minutes
Silence/Other	12.5%	5 minutes

What this means: In a class of 30, each student speaks for approximately 4 seconds per lesson.

Question Quality

Question Type	Per Lesson	Percentage
Closed-ended	12.8	87%
Open-ended	1.9	13%

Examples observed: "Is this correct?" / "What is 2+3?" / "Do you understand?"

Examples NOT observed: "Why do you think that happened?" / "How would you solve this differently?"

The Student Question Crisis

> **Key Insight:** In 522 observed lessons, students asked ZERO questions.

This indicates:

- No culture of inquiry
- Fear of appearing wrong

- No space created for questions
- Teachers not modeling curiosity

Part 7: AI vs. Human Coach Ratings

The 54-Point Gap

Indicator	AI Rating	Human Rating	Gap
Student Engagement	44%	98%	-54 pts
Students in Groups	44%	98%	-54 pts
Learning Objectives	84%	98%	-14 pts
Review Activity Done	82%	98%	-16 pts

Why Such Large Gaps?

Factor	Human Assessment	AI Assessment
Definition	"Students look attentive"	"Students ask questions, discuss"
Evidence	General impression	Specific transcript quotes
Threshold	Did it happen at all?	Did it happen meaningfully?
Focus	Programme compliance	Teaching quality

Conclusion: Both are measuring different things. Neither is "wrong," but they capture different aspects of classroom reality.

Part 8: AI Coach Limitations & Improvement Opportunities

What AI Is Missing

Field observations revealed several gaps in what the AI coach captures and how it provides feedback:

Gap	Description	Impact
YouTube videos	Teachers improvise by showing educational videos	AI only analyzes audio; misses visual instruction
Board work	Writing on the board not captured	Key teaching moments invisible to AI
Group activities	Physical arrangement of students not detected	AI scores low on collaboration it can't see
Visual aids	Charts, posters, manipulatives in use	Good practice goes unrecognized

> *"Some teachers were actually improvising and using YouTube videos and showing them in the classrooms, but our AI coach was not capturing that." — Field Observation*

Feedback Quality Issues

Teachers raised specific concerns about AI feedback quality:

Issue	Teacher Feedback
Generic feedback	Not specific enough to be actionable
English only	Feedback in English, not Urdu—limits accessibility
Repetitive criticism	Continuously told students aren't asking critical questions
No solutions provided	AI identifies problems but doesn't show HOW to fix them

The Student Question Dilemma

The AI consistently rated teachers low on student participation because students weren't asking questions. But teachers explained the deeper problem:

> *"They don't know how to do that because these kids are not used to asking questions. Even if they give them space, kids don't know how to ask those questions." — Teacher Perspective*

The Cultural Context:

- Pakistani classroom culture is traditionally teacher-centered
- Students are taught to listen, not question
- Asking questions can be seen as challenging the teacher
- Students lack the skills and confidence to formulate questions
- This is a systemic issue, not individual teacher failure

What This Reveals: The Missing Feedback Loop

Current State:

```
AI observes → AI gives feedback → Teacher receives → (No response mechanism)
```

What's Needed:

```
AI observes → AI gives feedback → Teacher responds with context → AI adjusts →
Better feedback
```

> **Key Insight:** There should be a conversation between AI feedback and teacher feedback so that the AI can get full perspective and give feedback accordingly.

Teacher input is essential because:

— AI doesn't see the full picture (no video, no board, no context)
— Teachers know their students' baseline capabilities
— Cultural context affects what's possible
— Teachers can explain WHY certain things didn't happen

Recommendations for AI Coach Improvement

Recommendation	Rationale
Add Urdu feedback	Make feedback accessible to all teachers
Include "how to" prompts	Don't just say "students should ask questions"—show teachers HOW to prime students
Enable photo uploads	Teachers upload photos of board work, group arrangements as evidence
Create feedback dialogue	Let teachers respond to AI feedback; AI adjusts based on context
Recognize improvisation	Credit teachers for using videos, visual aids (even if AI can't analyze them)
Provide question scaffolds	Give teachers specific prompts to train students in asking questions

Proposed: Student Question Training Module

Since students don't know HOW to ask questions, AI feedback should include:

Question Starters for Teachers to Model:

— "Can you explain WHY...?"
— "What would happen IF...?"
— "How is this different from...?"
— "I don't understand... can you show me?"

Priming Techniques:

— "Today, I want each group to come up with ONE question"
— "Before we finish, who has a 'I wonder...' question?"

— "Let's practice: What's something you're still confused about?"

The Shift Required:

— From: "Students should ask questions" (AI criticism)

— To: "Here's how to train students to ask questions" (AI coaching)

Part 9: Student Experience

Kids Enjoyed It

> *"Kids seemed to be enjoying it a lot." — Field Observation*

The hands-on materials, differentiated activities, and focused FLN instruction appear to engage students during the winter school programme.

But Will It Transfer?

Open questions:

- Will students remain engaged with larger class sizes?
- Will the novelty wear off in main season?
- Can the same activities work with 50+ students?
- How does curriculum pressure affect student experience?

Part 10: Scalability Concerns

Winter School vs. Main Season

Factor	Winter School	Main Season
Curriculum pressure	None	High
Class size	Smaller	50+ students
Teacher focus	FLN only	Full curriculum
Incentive	PKR 35,000	None
Time available	Protected	Competing demands
Teacher motivation	High (paid)	Variable

Teacher Concerns About Main Season

> *"These kinds of programmes focusing on FLN usually work during winter or summer vacations because teachers have more time." — Teacher Feedback*

Specific concerns:

- "During regular session, we're worried about completing curriculum"
- "Too many students to focus on foundational skills"
- "Attendance is a challenge"

Critical Questions for Scale

- Can we replicate compliance without PKR 35,000?
- Will teachers continue uploading without payment tied to it?
- How do we adapt FLN practices for 50+ student classrooms?
- Can indirect benefits persist without observation pressure?

— What's a realistic expectation for main season?

Part 11: Emerging Recommendations

For Immediate Action

Recommendation	Rationale
Maintain incentive structure	Compliance is high because of it; indirect benefits are real
Extend coach presence	Coaches are valued; essential for troubleshooting
Continue phonics reinforcement	One week isn't enough; teachers still confuse letters/sounds
Celebrate resource usage	Teachers ARE using pebbles, materials—build on this strength

For Main Season Pilot

Recommendation	Rationale
Test without incentive	See if indirect benefits persist
Start with willing teachers	Volunteers more likely to sustain practice
Adapt for large classes	Current plans assume small groups
Integrate with curriculum	FLN shouldn't compete; show how it supports

For Coach Training

Recommendation	Rationale
Shift role over time	Week 1-2: Tech support → Month 2+: Pedagogy coaching
Train on phonics gap	Coaches should identify letters-vs-sounds confusion

Calibrate with AI	54-point gap needs alignment

For Research

Recommendation	Rationale
Study incentive effect	What happens when payment stops?
Track phonics mastery	Pre/post assessments on teacher understanding
Validate AI thresholds	Is 44% too strict or is 98% too lenient?

For AI Coach Improvement (Priority)

Recommendation	Rationale
Add Urdu language feedback	Current English-only feedback limits accessibility
Include actionable "how to" prompts	Don't just criticize low student questions—show HOW to prime students
Enable photo evidence uploads	Teachers upload board work, group arrangements as proof
Create teacher response mechanism	Two-way dialogue so AI gets context before judging
Provide question scaffolds	Specific prompts teachers can use to train students to ask questions
Recognize visual instruction	Credit YouTube videos, visual aids (even if AI can't analyze them)
Adjust for cultural context	Pakistani students not trained to question—AI should account for baseline

The Feedback Loop Fix

Current Problem:

```
AI: "Students aren't asking critical questions" (repeated criticism)
Teacher: "I don't know how to make them do that" (no support)
Result: Frustration, no improvement
```

Proposed Solution:

```
AI: "Students aren't asking questions yet. Here are 3 prompts to try tomorrow:
     1. 'What's one thing you're still confused about?'
     2. 'Can anyone think of a different way to solve this?'
     3. 'What would happen if we changed X to Y?'
     Try these and upload your next recording!"
Teacher: Tries prompts, uploads new recording
AI: "Great! I heard 2 student questions today. Keep going!"
Result: Gradual improvement with guidance
```

Part 12: Open Questions for Team Discussion

---

1. **Incentive Dependency:** How do we transition from extrinsic (payment) to intrinsic (professional pride) motivation?

2. **Main Season Reality:** Is it realistic to expect FLN practices in 50+ student classrooms with full curriculum pressure?

3. **Phonics Mastery:** What additional support do teachers need to truly understand and implement phonics (not just know it matters)?

4. **Observation Sustainability:** Will teachers continue uploading observations when it's not tied to payment?

5. **Human Coach Capacity:** Can we scale human coach support, or do we need to shift more to AI coaching?

6. **Success Definition:** What does "success" look like for the main season? Same compliance? Or different targets?

7. **Student Outcomes:** How do we link teaching quality to actual student learning gains?

8. **AI Feedback Quality:** How do we make AI feedback actionable, not just critical? How do we enable two-way dialogue?

9. **Cultural Adaptation:** How should AI adjust expectations for contexts where students aren't culturally trained to ask questions?

10. **Evidence Beyond Audio:** Should we add photo uploads, video clips, or other evidence to give AI fuller picture?

Appendix: Data Summary

Observation Statistics

```
Total Observations: 576
├── AI Observations: 522
│   ├── With complete scores: 200 (analyzed)
│   ├── With talk time data: 522
│   └── With question counts: 522
└── Human Observations: 54
    ├── Yes/No indicators: ~15
    └── Rating scales: ~5

Teachers with Both AI and Human: 34
Lesson Plans Analyzed: 191
```

Implementation Gap Summary

Practice	In Plans	Observed	Gap
Open-Ended Questions	100%	13%	-87%
Student Collaboration	97%	6%	-91%
Real-Life Connections	100%	16%	-84%
Student Questions	16%	0%	-16%

Talk Time Distribution

Speaker	% of Time	Minutes (40-min)
Teacher	81.8%	33 min
Students	5.7%	2 min
Silence/Other	12.5%	5 min

Programme Incentive Structure

Element	Detail
Payment per teacher	PKR 35,000 (~$125)
Condition 1	Complete all lesson plans
Condition 2	Upload all observations
Condition 3	Attend training
Training duration	One week

Conclusion

The Balochistan Winter School Programme demonstrates what's possible when teachers have protected time, adequate support, and clear incentives. The results are promising: high compliance, positive attitudes, real (if indirect) improvements in classroom behaviour.

But sustainability remains an open question. The programme's success may depend heavily on the PKR 35,000 incentive and the protected winter school context. Scaling to the main season—with curriculum pressure, larger classes, and no payment—will require adaptation.

The key insight from field observations:

> *"One week of training can get them started, but continuous reinforcers are important. Once teachers get the hang of it, only then can they pass on the knowledge."*

Teachers know phonics matters. They know collaboration is good. They know open-ended questions promote deeper thinking. **But knowing is not doing.** The implementation gap persists.

A second key insight emerged about AI coaching:

> *"The AI was giving very generic feedback, and it was in English, not in Urdu. Teachers don't know HOW to make students ask questions—they need prompts and techniques, not just criticism."*

AI feedback must evolve from critic to coach. Identifying problems isn't enough. Teachers need actionable guidance, cultural context awareness, and a two-way dialogue where their input shapes the feedback.

The path forward requires:

- Continued incentives (at least initially)
- Ongoing coaching (not just initial training)
- Realistic expectations (main season will be harder)
- Hybrid AI-human support (frequency + depth)
- AI coach evolution (from critic to coach—with Urdu feedback, actionable prompts, and two-way dialogue)
- Photo evidence system (capture what audio misses—board work, groups, visual aids)
- Patience (behaviour change takes time)

Report Status: Draft for team review

Next Steps: Add team input, validate AI coach recommendations, plan main season pilot

Field Intelligence Report combining qualitative observations and quantitative data analysis

Balochistan Winter School Programme, January 2026