EDTECH HUB AI OBSERVATORY AND ACTION LAB

Teachers-in-the-Lead Toolkit

A toolkit for centering educators in the design and implementation of AI-enabled tools

AI Observatory

EdTech Hub
Clear evidence, better decisions, more learning.

UK International Development
Partnership | Progress | Prosperity

Contents

1. Purpose of this Toolkit

The problem we are trying to solve

Artificial Intelligence (AI) tools for education are being built at a rapid pace, and often without teachers in the room. While there is a flourish of information increasingly available about EdTech and AI tools (what they "can and can't do") there is little discussion about the practicalities of bringing teachers into the design, testing, and implementation processes that drive the development of such tools.

The cost of getting this wrong is not hypothetical. The history of EdTech includes a myriad of well-funded tools that never made it into sustained classroom use and never moved learning. With AI these tools are easier than ever to develop and disseminate across education systems. South Korea offers a cautionary example at national scale. A USD 70 million initiative to deploy AI-powered digital textbooks aimed to reach all schools by 2025 and was adopted by fewer than 30% of them. A survey found 87.4% of teachers felt unprepared to use AI and wanted greater input and choice. The technology itself was not the binding constraint to adoption (↕Sheffey, 2025).

Without teacher input, decisions about what gets built and how it works are made only by developers, funders, and policymakers - most of whom are not in classrooms, and (as often happens in low- and middle-income countries) many of whom are not in the country. Students bear the cost and the digital divide widens.The International Task Force on Teachers for Education 2030 is clear on this point: AI's benefits in education largely depend upon the expertise and ethical judgement of teachers (↕Teacher Task Force, 2025). Without them, effective implementation is simply not possible.

But how should those building EdTech tools in the age of AI put teachers in the lead, and ground their approach in the emerging evidence? Building on EdTech Hub's Innovation practice, as well as evidence from its AI Observatory and Action Lab, including a portfolio of Teachers-in-the-Lead Sandboxes, this Toolkit aims to provide a practical guide for decision-makers in low- and middle-income countries to answer that question.

Who this toolkit is for

Three groups shape what a teacher-facing AI tool becomes in low- and middle-income countries. Each holds a different set of decisions at different points in the development, testing, and roll-out of the tool.

All three are core audiences for this toolkit:

- **EdTech product developers and technical teams build the tool.** They hold the decisions that determine whether it fits a teacher's day: what the interface assumes, which platform it lives on, which languages it works in, which underlying model it works on, and what happens to a teacher's feedback once it arrives.
- **Implementing partners and programme teams put the tool in front of teachers.** They sit between the developer and the school; convening teachers, holding the relationships that make honest feedback possible, and running testing in real classrooms. They are usually the only group positioned to make teacher engagement structural rather than occasional, because they control the rhythm of contact across a design cycle.
- **Ministry of education officials set the conditions and govern roll-out.** They determine what a procurement requires, what gets funded, what a vendor must demonstrate, and which policies help or hinder classroom use. The decisions that determine whether a tool survives beyond a pilot - institutional ownership, recurring cost, and investment in local language capability - sit at this level.

What you can do with it

This toolkit gives teams methods to pick up, try in their context, and adapt in order to leverage AI tools to support teachers. Used well, it should help you to better define how you expect AI to support teachers you work with, provide some evidence-based guidance on which AI use cases are most promising to pursue, and provide a menu of options on how to meaningfully engage teachers, gather and act on their feedback, and avoid the design pitfalls that have sunk a generation of edtech.

The three steps

The toolkit has three steps, each a prompt for reflection. They're designed to fit whatever you're actually working on, so start with whichever feels most pressing.

Three-step process

1

Build your hypothesis

What are you trying to change, for which teachers, and how would you know if it worked?

2

Learn from the evidence

What do we already know about this AI use case for teachers, and how does this shape your approach?

3

Put teachers in the lead

Which methods will get teachers into the design, testing, and development of the AI solution itself?

Considering opportunity cost, and power

Every method in this toolkit has a cost. When teachers attend a design sprint, complete a survey, or join a co-design session, they are spending time they could have spent teaching, planning, or resting. In contexts where teachers are already carrying heavy workloads, that trade-off is real.

This does not mean teacher engagement is not worth it. It means the obligation runs both ways. If you are asking for teachers' time, the output of that time must visibly improve the tool they use.

Every method in this toolkit also puts a developer or researcher in a room with a teacher, and that is not a relationship between equals. The developer holds the tool, the funding, and the decision about what gets built. The teacher is often being asked for their time by someone more senior, better resourced, or simply less exposed to the risk if the tool fails.

Design Justice, the set of principles developed by a network of designers and community organisers, argues that a design process should centre the voices of those most affected by it and put their interests ahead of the designers' (↑Costanza-Chock & Braman, 2020). Liberatory Design, developed at Stanford's d.school, pushes further: it treats redistributing power between those who hold it and those affected by a design as the point of the process, not a side effect of doing it well (↑Anaissie et al., 2026). Both exist because good intentions do not close a power gap on their own.

Before running any method, ask yourself, who holds the power in the room, what decision will this change, who will hear what we learnt, and what will teachers get back.

Step 1.
Build your hypothesis

Everyone is asking what AI can do for teachers. The harder question is what can teachers do with AI.

This step adapts the AI Observatory's hypothesis-based approach for teams building teacher-facing tools. Work through it in a team, or use it to pressure-test thinking you have already done. The output is a hypothesis you can actually test.

1.1 Define your vision

What are you trying to achieve? Before reaching for tools or strategies, name what you are working toward and for whom. A vision worth working towards names an audience and connects an AI use case to a concrete change in their experience.

THINK ABOUT

What role might AI play for the teachers you work with? What does their working week look like if this goes well? Be specific about which teachers: a Grade 3 teacher in a multi-grade rural classroom and a secondary science teacher in a well-connected city school do not have the same problems.

EXAMPLE

> **"I want the teachers I work with to leverage AI tools to gain efficiencies, and ultimately improve their instruction."**
>
> Notice that this names a specific audience (teachers) and links AI to a concrete outcome (efficiency leading to better instruction). Now sharpen it further: which teachers, teaching what, under what conditions?

1.2 Spot your barriers

What is actually in the way? The more entrenched your barriers, the sharper your hypothesis. Identify two to four core challenges in your context before jumping to solutions, considering people and capacity, resources and access, systems and structures, connectivity and infrastructure, and awareness and motivation.

EXAMPLE BARRIERS

- Low AI literacy: teachers do not know what tools exist or how to use them.
- Limited time and flexibility in the school day to try new approaches.
- Unclear value proposition: teachers do not see how AI improves their work.
- Connectivity and resource gaps in under-served and rural areas.

- Policy friction: for example, a requirement for hard-copy lesson plans that forces teachers to transcribe AI-generated output by hand.

WHAT TEACHERS TOLD US

EdTech Hub surveyed over 800 teachers across 29 low- and middle-income countries before launching our Teachers-in-the-Lead Sandboxes. Four unmet needs emerged.

1. Teachers wanted AI to guide better pedagogical approaches, not just save them time in routine tasks.
2. They wanted frameworks for thinking critically about how AI affects student cognition.
3. They wanted formal guidance so they could explore AI safely.
4. They wanted accessible, teacher-led training, especially for those who feel left behind.

The same survey found that only 45% of participants had reliable access to electricity, internet, and devices at school - and that 82% were already using AI anyway. Teachers do not wait for perfect conditions. Your barriers list should reflect the conditions they actually work in, not the ones your tool assumes.

1.3 Build your hypothesis

Turn your thinking into something testable. A good hypothesis connects an action to a change, and is specific enough that you would know whether it worked. The if-then format helps.

> **THE FORMULA**
>
> If [action that addresses a barrier], then [expected outcome for teachers or learners].

THE HYPOTHESIS TESTED IN OUR TEACHERS-IN-THE-LEAD SANDBOXES

If teachers are meaningfully engaged in shaping how AI is used, we will unlock more effective, equitable, and grounded approaches that improve teacher well-being, enhance the quality of teaching, and strengthen student learning outcomes.

AND HOW EACH PARTNER MADE IT THEIR OWN

Taleemabad, Pakistan	If teachers co-design AI tools with developers, rather than having tools designed for them, might we see higher uptake rates and improved outcomes?
Rising Academies, Rwanda	If teachers have access to an instructional support tool that reduces planning time, will they spend more time on instruction and improve teaching quality?
Aga Khan Foundation, Tanzania	How might we use AI to strengthen peer-to-peer support and informal learning networks to improve the quality of professional development for all teachers?

EXPECT YOUR HYPOTHESIS TO SHIFT THE MORE YOU LEARN FROM TEACHERS

In all three sandboxes it did — and in each case the shift came from teachers.

- **Taleemabad** began by assuming most teachers worked in single-grade classrooms. A month in, schools revealed teachers were managing up to five grades at once, and multi-grade planning became the central design challenge.
- **Rising Academies** expected Tari to be used for curriculum-aligned lesson planning. Teachers predominantly used it for English language support, including privately rehearsing pronunciation before students arrived.
- **The Aga Khan Foundation** framed its sandbox around strengthening peer-to-peer learning. All four teacher teams converged instead on lesson preparation, and on spreading AI literacy to colleagues.

None of these were failures. Each was a hypothesis meeting reality and getting sharper, which is exactly what the next two steps are designed to make happen — early, cheaply, and before the build is locked.

A downloadable card deck is also available from the AI Observatory, which walks through the same steps towards hypothesis development, with guided questions and examples.

Your hypothesis is ready to move on when:

- It names specific teachers, in specific conditions, not "teachers".
- You can point to the barrier it addresses, and say how you know that barrier is real.
- You could describe a result that would tell you it was wrong.
- You know which use case it belongs to, so you know which evidence to read first.

Take it to Step 2 to see what the field already knows. Then take it to Step 3 and let teachers tell you the rest.

Step 2.
Learn from the evidence

You are almost certainly not the first team to try this use case. Learn from others, and give back what you find.

Through our Learning Brief Series, EdTech Hub has assessed the evidence on teacher-AI use cases and offers a helpful way to organise this landscape, through three horizons:

1. **Upgrade**, where AI helps teachers to complete existing tasks more efficiently;
2. **Disrupt**, where it changes the underlying practices of teaching and redistributes how teachers spend their time; and
3. **Transform**, where the very nature of the teacher's role is rethought to account for AI's presence in education.

Consider how the existing evidence might shape your own testing with teachers. Take your hypothesis and ask yourself:

- Which use case does my hypothesis actually sit in, and which horizon?
- What has worked elsewhere, in conditions like mine — and what has consistently failed?
- What does the evidence say to expect, so that I am not claiming more than the field can currently support?
- Where is the evidence thin? That is not a reason to avoid the use case, but it should change how carefully you test it.

The table below provides a summary of AI use cases for teachers across Three Horizons from our series of Learning Briefs on the topic.

Three Horizons	Horizon 1 Current dominant system where change happens incrementally within existing structures	Horizon 2 Turbulent space of innovation & experimentation as society shifts from the status quo to a new paradigm	Horizon 3 Radical new visions of the future that exist in the fringes today compete to be the future dominant system
What does this mean for empowering teachers?	**Upgrade:** Early integration of AI into teachers' roles, where some existing tasks and processes become streamlined or enhanced, while new demands emerge around oversight, adaptation, and judgement.	**Disrupt:** AI use becomes more formally evaluated and embedded in teaching, addressing some early identified issues, and disrupting aspects of teachers' roles in ways that shift their practice or time allocation.	**Transform:** The role of teachers is fundamentally changed, with AI being a core facet of the teaching and learning experience in novel ways.
Use cases	AI for content creation and lesson planning AI for administrative tasks AI for assessments, grading, and feedback AI for professional development	AI-powered digital personalised learning AI tutor chatbots AI teaching assistants AI-powered inclusive education and language accessibility	"Teacherless" classrooms Teacher-AI complementarity and hybrid intelligence Multi-intelligence learning in unbound learning environments

Below we provide summaries of the evidence across these three horizons, as captured in our Learning Brief series, which can be used to check your hypothesis, and consider what kinds of AI tools may be helpful.

2.1 Horizon 1: Upgrade

Early integration of AI into teachers' existing roles, where some tasks become streamlined or enhanced. This is where most teachers are, where the evidence is strongest.

CONTENT CREATION AND LESSON PLANNING

What it is. AI generates curriculum-aligned lesson plans, worksheets, quizzes, and question banks, and adapts existing materials to local language, culture, and learner level.

What the evidence says. This is among the most promising use cases for AI supporting teachers. A randomised controlled trial in England found science teachers using ChatGPT with a guide cut planning time by around 31%, though usage decreased as the trial progressed (↑Roy et al., 2024). In Pakistan, Lesson Genie cut planning from roughly an hour to 1 or 2 minutes (↑Burgess, 2026) . Country-contextualised, curriculum-aligned platforms tend to outperform raw Large Language Model (LLM) use, because alignment and guardrails are handled for the teacher rather than by them. While promising evidence if mixed on the ultimate impact of these tools on learning outcomes. One recent study has also suggested that AI generated material development by teachers can lead to cognitive offloading and worse outcomes for learners (↑Sungu et al., 2026).

> **WATCH OUT**
> Untailored LLM output may not align with national curricula, assessment standards, or exams, especially when drawing on high-income curricular frameworks. Pedagogical quality depends heavily on prompting, which assumes teacher AI literacy you may not have.

ADMINISTRATIVE TASKS, CLASSROOM MANAGEMENT, AND DATA-DRIVEN INSIGHTS

What it is. Automating attendance, timetabling, parent communication, data entry, and reporting, and surfacing insights from school data.

What the evidence says. Evidence in low- and middle-income countries is currently inconclusive, and there are few tools designed for these contexts. Generic tools for tasks like emailing parents are often not relevant, and assume a digitally confident teacher with a device. This use case has a gap between how much attention it gets in pitches versus how much evidence sits behind it.

> **WATCH OUT**
>
> AI systems make their own errors, and correcting them takes teacher time, undermining the efficiency gain that justified the tool. Tasks that look educationally unproductive are often not — taking a register (attendance) is also how a teacher learns names and notices who is absent. Privacy risks also surface relevant to this use case: for instance, facial recognition approaches for attendance have drawn significant pushback on privacy and safeguarding grounds and should be approached with real caution.

ASSESSMENT, GRADING, AND FEEDBACK

What it is. AI assessing reading fluency, grading written work, and generating feedback for learners.

What the evidence says. The evidence is so far mixed, split sharply by assessment type. For reading assessment, evidence from India and South Africa suggests AI can be quite effective at assessing fluency and has adapted well to local languages after training. For essay grading with LLMs, results are mixed and depend heavily on the prompts, rubrics, and benchmark data provided.

> **WATCH OUT**
>
> Data privacy and safeguarding risks are material where student work is involved. Grading tools can struggle with children's voices and nuanced writing, and essays that do not follow AI conventions may be downgraded. Students have shown mistrust of AI-only feedback. Mitigating bias often means teachers supplying rubrics and annotated examples, which adds workload rather than removing it.

PROFESSIONAL DEVELOPMENT AND COACHING

What it is. AI-powered coaching, on-demand instructional support, simulated practice environments, and continuing professional development delivered through chat.

What the evidence says. The examples show real promise for both in-service and pre-service teachers, and this is where two of the three Teachers-in-the-Lead Sandboxes sat. The features that consistently improve quality are: tailoring to the local context and curriculum, keeping data costs low or working offline, delivering through a platform teachers already use, and building teachers' AI literacy alongside the tool.

> **WATCH OUT**
> Polished answers are not necessarily accurate ones, and they may be perceived as more reliable than they are — which can reduce the exercise of professional judgement. Studies of remote coaching show impact declining over time as accountability and support fall away. AI can supplement professional development effectively; substituting it wholesale for in-person support removes the relational elements that make professional development work.

2.2 Horizon 2: Disrupt

Under this Horizon, AI use becomes more formally evaluated and embedded, disrupting aspects of teachers' roles in ways that shift their practices or time allocations.

AI-ENHANCED DIGITAL PERSONALISED LEARNING (DPL) PLATFORMS

What it is. Adaptive platforms that diagnose learner level and route learners through personalised content pathways, with the teacher assigning tasks and interpreting dashboards.

What the evidence says. Evidence from Kenya and India indicates that AI-enhanced personalised learning can generate moderate learning gains at relatively low cost - but most effective examples all embed active teacher involvement, curriculum alignment, low-cost reliable hardware, and real-time feedback loops that let teachers target support. The teacher is part of the mechanism, not a bystander to it (↑Adam & Lester, 2025).

WATCH OUT

Rigorous evidence remains limited relative to the number of DPL platforms on the market. Many report internal engagement metrics with no independent evaluation of learning outcomes, and many now market themselves as AI-enhanced without clarifying how AI is integrated. Personalised learning was already promising before GenAI, so it is often unclear whether gains come from the AI layer or from foundational adaptive and teaching-at-the-right-level design.

AI TUTOR CHATBOTS

What it is. Conversational tutors for learners, ranging from generic LLM interfaces to highly contextualised subject-specific tutors.

What the evidence says. The evidence of AI-based tutors improving specific learner outcomes is nascent, but promising. WhatsApp-delivered, low-bandwidth designs such as Rori in Ghana have produced substantial improvements in learner outcomes (↕Henkel et al., 2024) and proven more robust and scalable in low-resource contexts than approaches requiring account setup and steady connectivity . However, it is frequently unclear whether observed gains come from the AI, from increased practice time, from the structured pedagogy embedded in the tool, or from novelty.

WATCH OUT

One study found that when access to OpenAI's GPT-4 model was withdrawn, students performed 17% worse than those who never had access - evidence that these tools can become a crutch (↕Bastani et al., 2024). Related concerns include cognitive offloading and its effect on critical thinking. Costly humanoid or anthropomorphic tutors warrant particular scrutiny, both for weak evidence and for the teacher-replacement narrative they reinforce.

AI TEACHING ASSISTANTS

What it is. Integrated assistants supporting the range of teacher's task-planning, materials development, pedagogy questions, and assessment - usually through a single conversational interface.

What the evidence says. Causal evidence is not yet available, but the examples are promising. Contextual adaptation is the recurring theme in the assistants that work: alignment to a national curriculum, delivery

through a low-bandwidth channel teachers are already on, and buy-in from the institutions teachers trust. Worth noting, time saving is not automatic: teachers often spend hours repairing and reworking outputs they find inadequate, which is hidden labour that industry workload claims tend not to account for.

> **WATCH OUT**
> Effectiveness depends far more on integration into teacher workflows, training, local relevance, and infrastructure than on the model itself. And more is not necessarily better: where teams were offered several overlapping AI platforms, teachers reported feeling overwhelmed and experienced decision paralysis. Professional development focused on a few features that solve a specific problem is often more effective.

AI-POWERED INCLUSIVE EDUCATION AND LANGUAGE ACCESSIBILITY

What it is. Reading support and assessment, translation, speech and text technologies, gesture recognition, and tools supporting learners with disabilities.

What the evidence says. This is a genuinely promising area with a growing evidence base on product accuracy, particularly for reading assistance and assessment, and for minority and sign languages. It is also directly relevant to the language problem our Teachers-in-the-Lead sandboxes surfaced: teachers across many contexts are expected to teach in a language other than their primary.

> **WATCH OUT**
> Geographic, linguistic, and cultural bias in AI tools are the central risks: models trained on the "average" user embed assumptions that force users to adapt to the technology. Poor performance in local languages is a central equity concern. EdTech Hub's partners Fab-AI are currently benchmarking AI tools for multi-lingual education use to bring more transparency to this challenge.

2.3 Horizon 3: Transform

Radical visions in which the teacher's role is fundamentally changed. This horizon is more speculative than evidence-based, and we present it as trends and futures rather than findings. It is included because the assumptions embedded in transform-horizon products are already shaping what gets built at the upgrade horizon.

AI-LED LEARNING ENVIRONMENTS

What it is. Models where AI provides the instructional core - content, sequencing, diagnosis, feedback, assessment — and the human role shifts to supervision, motivation, and pastoral support.

What the evidence says. Three distinct cost models exist: a high-cost private school model in the United States, a mid-cost franchised tutoring model in China, and a low-cost government partnership model in India. Teachers' roles differ sharply between them, from technical supervisor to comprehensive guide. These offer some promising internal results on improvement of learning outcomes, but also significant concerns from those who argue they may provide too narrow a definition of learning, and in the long-term, hinder holistic development of learners. (↕Adam, 2026)

> **WATCH OUT**
> The human support required to make these models a high-quality holistic experience is itself costly, and unlikely to scale without quality loss. Displacing teachers from the instructional core has unexamined effects on learner motivation, identity, and relational engagement. And outsourcing the instructional core to a vendor raises real questions about governance, vendor lock-in, and who owns learner data.

TEACHER–AI COMPLEMENTARITY AND HYBRID INTELLIGENCE

What it is. Configurations in which teacher and AI work together in the classroom, from AI drafting and teacher reviewing, through to genuinely two-way collaboration.

What the evidence says. The term is used very loosely, which is itself worth knowing when reading vendor claims — it covers everything from a teacher editing an AI-drafted lesson plan to live support during teaching, and several frameworks now try to distinguish the levels. Genuine two-way collaboration, where teacher and AI practices shape each other, does not

yet seem to exist in practice. What the examples that do work have in common is that they preserve the teacher's judgement rather than routing around it: the strongest results come from designs where teachers can override what the system suggests. Complementarity is something a development team has to design for rather than something that follows from simply putting teachers and tools together (↑Adam, 2026).

> **WATCH OUT**
> Effective collaboration appears to require high teacher agency and well-developed metacognition, ethical reasoning, and judgement about when not to use AI. Primary and secondary contexts in low- and middle-income countries are almost entirely absent from this literature.

MULTI-INTELLIGENCE LEARNING ECOLOGY

What it is. A deliberately plural approach in which AI-led, teacher–AI, and human-only pathways each have a defined place, including protected AI-free spaces.

What the evidence says. This is the most speculative use case in the series and the least evidenced. It is included because it names something the other categories obscure: the decision about where not to use AI is a design decision, and currently an unmade one in most products. Human-only approaches — role modelling, mentoring, building perseverance, ethical judgement, genuine care - are not residual leftovers but part of teaching that carries significant weight.

> **WATCH OUT**
> Treating this as a design consideration rather than a philosophical one requires you to state, explicitly, which parts of the teacher's role your tool is deliberately staying out of. Very few products do. If yours does not, a teacher will make that decision for you, and you will find out about it in the usage data.

2.4 What to take from the evidence base

No use case here is a sure thing. Therefore, in considering adoption, it is important to weigh how much to rely on existing evidence versus test and collect data on your own implementation along the way. Here are a few recommendations on how to use these insights:

1. **Match your evidence burden to your ambition.** Consider how you might build on existing evidence and test a use case for yourself. Before you commit, consider what key assumptions you would like to test and whether these seem reasonable with your existing capacity, resources, and scale. Right size your own data collection and experimentation accordingly.
2. **Consider alternative, proven interventions.** Existing evidence-backed approaches to improving teacher performance and learner outcomes, including continuous professional development, structured pedagogy, teaching at the right level, and (some) digital personalised learning already exist and do not require AI. When considering AI use cases, weigh these alternatives, their respective costs, and sustainability.
3. **Cost the time it takes to verify AI output before you claim the time saving.** Most efficiency-based AI use cases require some amount of effort and time from a human-in-the-loop (need to verify, correct, or curated output, for instance). Consider the net benefit of time saved and check the task you are automating is one teachers actually spend time on.
4. **Choose fit over sophistication.** Tools anchored to national curriculum and existing teaching materials can remove work; general-purpose tools add a verification step. Tools delivered on a channel teachers already use can often allow for more engagement than tools that require onboarding. And weak local-language performance is a hard barrier to adoption, not an issue to address later. Consider these dimensions as crucial to AI tool uptake by teachers. Right-fit may ultimately mean simpler: smaller, specialised language models running on modest devices could, in the future, address several challenges to implementation in low- and middle-income contexts, such as offline capability, no recurring inference cost, and local language as the starting point rather than a translation layer.

Step 3.
Put teachers in the lead

Teachers can and should guide how AI is used to support them. But how you build tools with teachers rather than for them?

3.1 From “in the loop” to “in the lead”

The phrase “human in the loop” is borrowed from engineering, and in education it is a dangerous import. It positions AI as the primary actor and the teacher as a secondary safeguard. Pedagogy-first design inverts that: first you define what good teaching and learning look like, and then decide if and how AI fits.

In practice, being “in the lead” rather than “in the loop” shows up as a shift in the role teachers play. Across the Teachers-in-the-Lead sandboxes (↕Burgess, 2026) moved teachers reliably from subject to stakeholder and began to show what Product owners could look like.

Subject	**Tester**	**Stakeholder**	**Product owner**
Teachers are studied. You learn about them.	Teachers react to what you built. You learn whether it works.	Teachers shape the backlog. You learn what to build next.	Teachers direct the tool. They decide what it is for.

3.2 The five stages of teacher involvement

The twelve methods that follow are grouped into five stages in tool development where teachers' input can be sought. The phases are roughly sequential, but the loop between Iterate and Validate is where most testing happens, and teams frequently return to Understand when the data surprises them.

Understand:

Grasping teacher reality before building anything

Teachers show design teams what their working reality actually looks like, surfacing constraints, workarounds, and priorities that a team working from assumptions alone would miss entirely. This is where teachers' lived experiences, rather than a team's prior beliefs, should set the direction of the hypothesis before a single line of code is written.

Generate:

Creating ideas with teachers

Teachers surface and shape the ideas worth pursuing, drawing on classroom realities and priorities that a design team working alone would never think to ask about. Their input at this stage becomes a record developers can trace directly back to what a teacher said and why, rather than a list of features nobody asked for.

Iterate:

Co-designing and refining

As the tool takes shape, teachers feed a continuous stream of solicited and unsolicited feedback into every development sprint, telling the team what is working, what isn't, and what they're actually using the tool for. This ongoing signal - not a one-off consultation - is what changes the defined revisions at each cycle, and it's often teachers who reveal that a workflow assumption baked into the tool was wrong.

Validate:

Affirming and making sense a tools performance

When usage data shows an unexpected pattern or assumptions about how a tool is used need to be checked, teachers are the ones who can provide insight. Their explanation is what tells a team whether to build on a pattern or question it. It can also validate a model that's working well and catalyze its move to the next level.

Champion:

Building ownership and spreading adoption

Teachers who helped shape the tool become the ones who introduce it to colleagues, giving it a credibility and reach that a top-down rollout cannot replicate. Their willingness to keep using it unprompted is the clearest signal a team gets that the tool is genuinely useful, not just adopted because they were asked to.

3.3 How to choose between methods

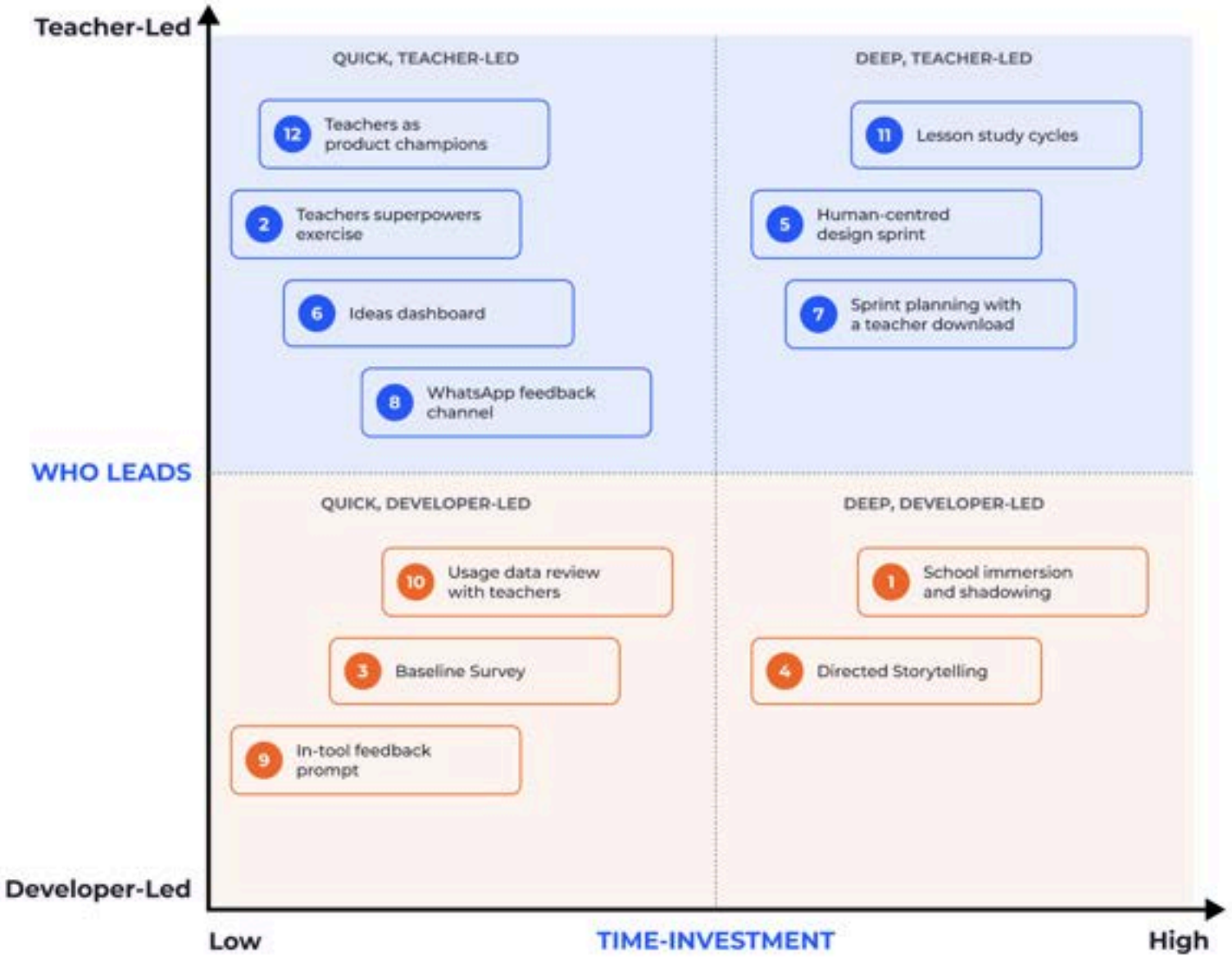
In this toolkit we share twelve methods for designing with teachers, at whichever stage you are at in your tool development. These are by no means the only approaches, and there are a number of excellent toolkits and playbooks for edtech co-design, which we encourage you to explore as well, such as IDEO's Field Guide to Human-Centered Design (↕IDEO, 2015) (a major inspiration for the methods captured here) as well as our EdTech Hub's own Sandbox Handbook (↕EdTech Hub, 2022).

In the matrix below, we have plotted these methods based on the time commitment required, and who leads. We encourage you to think carefully about how much time you can invest, and how you can truly put teachers in the lead without overburdening them. Treat that time as an investment, not a cost: an hour spent understanding what teachers need is an hour of engineering time you don't waste building something they won't use.

Two additional practical guidelines for choosing. If you have not yet built anything, spend your time in the Understand stage. And if your capacity is limited, prioritise a continuous method over a one-off method. For instance, a WhatsApp feedback channel keeps producing signals sprint after sprint, where even a good workshop gives you a single snapshot and then goes quiet.

Choosing a method: time-investment and who leads

The twelve methods in Step 3, plotted against the two things that decide whether you can actually run one.



Teachers lead the method The development team leads it

Methods 1-4 Understand . 5-6 Generate . 7-9 Iterate . 10-11 Validate . 12 Champion

Numbers refer to the method pages in Step 3 . Position within a quadrant is indictive, not measured

The Methods

Twelve ways to bring teachers further into the lead.

Understand

1. School immersion and shadowing

Send your team into schools to assist teachers directly. You will quickly find a problem worth solving, understand teacher led-innovations, and banish any assumptions you walked in with.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Before you build anything, or before a major pivot	**Developer / design team, hosted by teachers**	**2 to 4 weeks in schools**	**2 to 4 team members, 1 school cluster**

HERE'S HOW

1. Consider your pre-existing assumptions about your teachers' key challenges and how to solve them. Capture these prior to immersion to track what you learn, and how it's relevant to your tool's design.
2. Block time and budget for a true deep dive into teachers' experiences. Immersion is a method that can meaningfully shift a team's thinking regarding their tool's purpose and design, but it's among the most cost and time-intensive.
3. Agree with school leadership what your team will actually do. Shadowing from the back of the room gets you less insight than contributing directly. For example, teams might take the register, prepare a lesson, mark a set of books, and call a parent.
4. Record what you see and hear as concrete details and direct quotes, separately from your interpretation of it. Write down "she plans for three grades on one sheet of paper" before you write down "planning is hard".
5. Watch how teachers already improvise. The workarounds teachers have already invented are the strongest signal you will get about where a tool should live and what it should do.
6. At the end of each day, debrief as a team: what did you see, what surprised you, what assumptions did you get wrong. Assess how this shifts your tool's purpose, design, and implementation.

WATCH OUT

Immersion can feel expensive. If you can't fund weeks, be honest that you're trading depth for feasibility, and consider pairing a shorter visit with another Understand method to fill the gap.

CASE STUDY FROM OUR SANDBOXES

TALEEMABAD, PAKISTAN

Before building anything, Taleemabad's Learning Design Team spent a month embedded in schools. They observed, took attendance registers, planned lessons, and called parents. Two things emerged that reshaped the product. First, the team had assumed most teachers worked in single-grade classrooms; in fact teachers were managing two, three, or all five grades at once, and multi-grade planning became the central design challenge. Second, teachers showed the team they were already improvising with WhatsApp and ChatGPT, one of the reasons why Lesson Genie was built on WhatsApp rather than as a new app.

Understand

2. Teacher Superpowers exercise

Ask teachers what they would do if they had a superpower in the classroom and why. It's a simple way into what they actually need, without the conversation getting anchored to what AI can or can't do.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Early scoping, or to reopen a stuck conversation	**Teachers, lightly facilitated**	**45 to 60 minutes**	**5 to 15 teachers**

HERE'S HOW

1. Explain that the exercise is not about technology and there are no wrong answers. Do not mention AI, your product, or what is technically possible. The moment you do, teachers start editing themselves down to what they think you can build.
2. Ask each teacher to imagine they have been given a superpower they can use in their own classroom tomorrow. What is it? What does it let them see, know, or do that they cannot now?
3. Give people five minutes to write or draw alone before anyone speaks, so the first confident voice does not set the frame for the room.
4. Go around and have each teacher describe their superpower, prompting where needed. For instance, when in the last week did you most wish you had it?
5. Cluster the superpowers on a wall. Most will fall into a small number of families, for example: seeing what every learner understands, managing attention, or getting time back.
6. Write one sentence to describe each cluster and turn that insight statement into a ‘How might we’ question that you could answer in several different ways.

WATCH OUT

This method was originally used in the United States as a means of obtaining input from teachers that was not anchored on what is "technologically feasible" (↑Holstein & Aleven, 2022). The exercise translates, but the metaphor may not. In some contexts "superpower" reads as childish or unserious. Test the framing with two or three teachers first, and swap it for a local equivalent if it lands badly.

Understand

3. Baseline survey

A short structured survey that tells you what teachers have access to, what they already do with existing technology, and what they most want help with, before you design.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Before a design sprint, pilot, or feature scoping	**Developer / programme team**	**1 week to write, run, and read**	**25 to 100+ teachers**

HERE'S HOW

1. Before you draft your survey questions, decide what you'll do differently depending on each answer. If survey data gave you a definitive or surprising answer, would that insight meaningfully change your plan? If not, consider whether your question is worth asking.
2. A good starting point covers: what devices and connectivity do teachers actually have; what do they already use AI or digital tools for, if anything; which tasks take the most time; and what do they most want help with.
3. Ask about last week, not about general habits. “How many hours did you spend planning last week” gives you a usable number. “How much time do you spend planning” gives you an aspiration.
4. Keep it under 15 questions and run it in the language teachers think in, rather than defaulting to the official language of instruction.
5. Consider how the data impacts your design and implementation choices. Infrastructure answers are a priority. A tool that assumes reliable connectivity that is not there for most teachers will not lead to impact at scale. Share the headline findings back with the teachers who answered. It's a small effort that signals their time was worth something, which helps to build trust.

WATCH OUT

A survey tells you what teachers report, not what they do. Each of our Teachers-in-the-Lead sandboxes found a gap between the two. Use the survey to size and prioritise, then use immersion, school visits, or usage data to find out what is actually happening. Do not let a survey be your only contact with teachers before you build.

CASE STUDY FROM OUR SANDBOXES

ALL THREE TEACHERS-IN-THE-LEAD SANDBOXES

Rising Academies surveyed 40 teachers in Rwanda and immediately surfaced heavy lesson-preparation workloads and limited connectivity. The Aga Khan Foundation surveyed lead teachers in Tanzania before their design sprint, covering device and internet access, prior experience with AI, the tasks they most wanted help with, and their current professional development habits — and used the answers to set the workshop agenda.

Understand

4. Directed storytelling

Ask a teacher to walk you through one specific recent day. Specific stories carry details that general questions find hard to answer.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
When you need texture that a survey cannot give	**Developer / design team, with one teacher**	**60 to 90 minutes per teacher**	**1 teacher at a time, 5 to 8 in total**

HERE'S HOW

1. Probe your teacher for specific and focused insights into their practice: choose the moment you want the story about, and make it concrete and recent. Ask the teacher to start at the beginning and narrate forward in time. Where were you, what time was it, what did you do first.
2. Listen actively — Your job is to keep the story moving with “and then what?”, “what happened next?”, “who else was there?”. This will help give you the level of detail you need to identify key challenges or opportunities for intervention.
3. As much as possible, capture the story verbatim. The teacher’s chosen vocabulary is important, and combined with similar testimonies can be instructive about the words to use in your interface.
4. Capture stories from 5-10 teachers for a specific segment of their practices you’re trying to capture. Design your tool to respond to the identified challenges.
5. Over time, stories may lend themselves to be turned into themes or teacher personas. These can be defined as a short-hand for teacher inputs that designers and implementers should keep in mind.

WATCH OUT

This method depends on psychological safety. Without it, teachers give you the tidy, compliant version of the story — the one that reflects well on them, especially to someone from the organisation building the tool. Spend time building trust before you start, or better, have someone with an existing relationship lead the conversation.

CASE STUDY FROM THE SANDBOXES

RISING ACADEMIES, RWANDA

Rising Academies' school visits in Rwanda, where academic and product managers spent unstructured time listening to teachers. Those visits built the psychological safety a survey couldn't — teachers spoke openly about the difficulty of planning and teaching in English, and that insight directly reshaped Tari's design priorities.

Generate

5. Human-centred design sprint

Design sprints offer a method for rapid testing, learning, and iterating on specific aspects of a tool with its users (↑Google Ventures, 2019). It is an approach used widely in tech development, including in EdTech Hub Sandboxes. These often begin with a structured workshop that takes users (teachers) through generating ideas, building a prototype, and designing a way to test it in schools.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
When you have a question but not yet a solution	**Teachers, facilitated by the programme team**	**2-3 days workshop, weeks of testing as needed**	**20 to 30 teachers, in teams of 5 to 7**

HERE'S HOW

1. Run a sprint design workshop, which brings together a diverse pool of teachers eager to talk openly about their challenges and co-design solutions with your design team.
2. Invest time in building AI literacy as needed. Build a shared critical understanding of what AI is doing when it generates, what hallucination and bias look like in practice, and where it fails. As much as possible, demo AI tools and provide teachers with a direct user experience.
3. Move on to teachers' lived experiences. Map how things currently work in their classrooms, identify key challenges and root causes to effective instruction. Then, jump to solutions — teachers generate ideas individually or in groups, then cluster and vote on the most promising ideas to develop into prototypes. End with something tangible that a prototype a team can take back to their own school. This will allow you to build a minimum viable product to test with users, in small, rapid experiments. You can then test components of the tool as needed to inform design iterations.

4. Measure the user experience and impact of the tool. Consider both quantitative and qualitative data collection options to capture key performance indicators or test assumptions about how the tool might work.
5. Plan for a Sprint Review, which brings back together designers, users, and other relevant stakeholders to digest experiment data together. You can treat it as a second design workshop, where reflection and learning drive iteration.

WATCH OUT

Teachers are not product designers, and shouldn't be expected to become ones in an afternoon — they'll need real support to translate ideas into a testable prototype. Ensuring workshop facilitation guides them through the design thinking process is key.

CASE STUDY FROM THE SANDBOXES

AGA KHAN FOUNDATION, TANZANIA

The Aga Khan Foundation brought 28 lead teachers together for a three-day human-centred design workshop. These included AI exploration and tool ideation. Teachers left having built prototypes to test in their schools over eight weeks. A pre- and post-survey of 27 lead teachers found confidence in using AI rose from 15% to 63%, all 27 reported a clearer understanding of AI's capabilities and limits, and — importantly — the share who strongly agreed they understood AI's risks rose from 26% to 63%. This is a reminder that putting teachers in the lead is not only a design method, but a form of capacity building in its own right (↑Burgess, 2026).

Generate

6. Ideas dashboard

Capture group ideas using a common template, and then step back to look at common approaches and shared assumptions. Identify your most promising idea to prototype, and strengthen it with inputs from the others.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Immediately after any ideation session, before prototyping	**Teachers or product developers**	**≈30minutes**	**Teams of 4 to 7**

HERE'S HOW

1. Create a design team, each with the same one-page template to capture a design for an AI tool to test in response to your hypothesis. This should include what the idea is, who it is for, how it would work, and why it matters.
2. Prompt contributors to be highly specific. For example, "teachers" is too vague. "Lower-primary (P1–P3) teachers in double-shift schools responsible for teaching literacy" is better. The more precisely you define and understand the audience, the more your tool will be designed to fit their needs.
3. Have teams share their ideas with each other. Step back and make sure all completed dashboards are visible while others present their ideas. This is where cross-pollination happens, because you can combine features from different ideas.
4. Prioritise ideas to prototype with teachers directly or pass them to developers. Strengthen them by considering features or challenges identified in other ideas.
5. Test, learn, and iterate through design sprints, but always keep an eye on the ideas you passed on. They may offer important insights.

WATCH OUT

With school leaders in the room, teachers may converge on whichever idea seems safest to be seen agreeing with, rather than the one most worth testing. Build in individual reflection first, (e.g. dot voting) before open discussion, so people's honest reactions surface before the room settles on an answer.

CASE STUDY FROM THE SANDBOXES

AGA KHAN FOUNDATION, TANZANIA

On day three of our sandbox design sprint, teachers generated ideas individually, clustered and voted on the most promising, and then developed them using an Ideas Dashboard. This approach, taken from the Human Centred Design Toolkit (↕USAID & Aga Khan Foundation, 2025), covered what the idea was, who it was for, how it would work, and why it mattered. Four ideas emerged, each tackling a different barrier to peer learning. From there, teachers used Lovable (an AI-powered app-building tool) to actually build a working web app/prototype of their idea, rather than just describing it on paper (↕Burgess, 2026).

Iterate

7. Sprint planning with a teacher download

Start design and implementation sprints with a structured download of teacher feedback gathered, logged, and synthesised over the last sprint window and use it to prioritise the next iteration.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Every sprint, for the life of the product	**Product team, structurally accountable to teachers**	**15 to 30 minutes at the start of sprint planning**	**Whole product and academic team**

HERE'S HOW

1. Gather feedback from teachers into a “download”. Draw on multiple feedback streams across the sprint window. Feedback streams could include: a WhatsApp channel (method 8), an in-tool prompt (method 9), backend usage data (method 10), and notes from school visits.
2. Ground feedback with context. As much as possible, consider teachers’ profiles, including which school, grade, subject, and language. Without context, feedback is hard to act on. For example, "The plans are too long" means one thing from a Grade 1 teacher with 60 learners in a multi-grade room, and something different from a secondary teacher with a double length subject class.
3. Create space for product teams to look at feedback together in order to capture diverse perspectives. A product lead may focus on a feature request, whereas an academic lead may focus on a pedagogical problem. Without both lenses in the room, teams risk fixing symptoms (a complaint) rather than causes (a pedagogy issue). Involve teachers themselves as a counter reference if possible.
4. Cross-check what teachers say against what the data shows. Back end usage data will show what teachers actually did, not what they said they did. When there is a discrepancy, for example a well-rated feature nobody uses or a popular workflow nobody talks about, go back to school to learn more.

5. Open sprint planning with the download so that teacher feedback takes priority. Name the two or three high priority themes or challenges which your team will act on.
6. Close the feedback loop by telling teachers what features changed and when they will see them. This will encourage teachers to keep voicing their opinions. Then keep a feedback log across sprints so that you can spot patterns and recurring issues that might not show up immediately.

WATCH OUT

The "download" only works if it has teeth. If teacher feedback is read out and then the sprint proceeds exactly as already planned, teachers learn quickly that the ritual is performative, and the quality of feedback collapses.

CASE STUDY FROM THE SANDBOXES

RISING ACADEMIES, RWANDA

Every Tari sprint began with a structured download of what teachers had shared about using the tool, supported by weekly huddles, where the academic and product teams synthesised teacher insights. This was the architecture that kept teachers' voices audible through the product iteration cycle. For instance, when teachers said the American English accent on Tari was not relatable, the developers recorded one of the Rwandan teachers and made that the default voice in Rwanda and, as a result, Tari uptake increased.

Iterate

8. Messaging platform feedback channel

An open group on a platform teachers already use (WhatsApp or other), connecting them directly to the people building the tool. Unsolicited feedback is often the most honest kind.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
From first pilot onwards, continuously	**Shared — teachers post, the team responds**	**Ongoing, 1 hour of community management to run well**	**15 to 30 teachers per group**

HERE'S HOW

1. Use a digital platform teachers are already on (WhatsApp, Telegram, Facebook, are some examples). The goal is zero onboarding cost, so do not create a new channel, forum, or app for this.
2. Put product developers in the chat group too, alongside a community manager. The value in this inclusion is creating the shortest path between what teachers notice and the person who can change it.
3. The group needs someone teachers already know and trust running it week-to-week. This may be a teacher coach, an academic programme lead, or a lead teacher from within the group itself. Without someone actively present, asking questions and responding to what comes in, it's hard to nurture feedback.
4. Have the moderator respond as much as possible to acknowledge concerns and feedback. This includes providing honest answers around constraints. If teachers can't distinguish between being ignored and being constrained by a cost, a technical limit, or a policy requirement, they have no way to calibrate whether the group is worth their time.
5. Capture what is expressed in the community. Screenshot and log exchanges, pictures, and shared insights. Follow up as needed to identify more details or understand a point more clearly. You should

review and thematically analyse this information to inform your design sprints.

6. When a teacher's message shifts design or implementation, be sure to credit the teacher or make clear the idea came directly from someone in the group. Visible influence is what turns a feedback channel into an ownership channel. This demonstrates that teachers are shaping the tool, not just being consulted, and they are likely to champion tools where they can point to a feature and say, "I asked for this".

WATCH OUT

Power dynamics are crucial. If head teachers and classroom teachers are in the same group, you may hear a filtered version of classroom reality. Consider separate groups, and pay attention to who never posts. And be clear with yourself that a WhatsApp group is a supplement to being in schools, not a substitute for it.

CASE STUDY FROM OUR SANDBOXES

TALEEMABAD, PAKISTAN AND RISING ACADEMIES, RWANDA

Taleemabad ran a WhatsApp onboarding group with 17 head teachers that connected them directly to the design team and academic lead, capturing unsolicited, real-time responses to Lesson Genie, including screenshots of lesson plans teachers had actually used. Rising Academies ran WhatsApp groups connecting teachers directly with academic teams and school leaders. In Tanzania, WhatsApp groups for lead teachers kept a continuous record of what was happening in schools between formal check-ins (↑Burgess, 2026).

Iterate

9. In-tool feedback prompt

A short rating and comment prompt triggered inside the tool at the end of a task, while the details about the user experience are still clear.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Once the tool has regular users	**Developer, answered by teachers**	**15 seconds per teacher, per use**	**Every user**

HERE'S HOW

1. Trigger the prompt at the natural end of a task, not on opening or closing the tool. For example, ask after the lesson plan is generated, not before.
2. Keep the data input requested from teachers to one rating and one optional open question. Teachers will engage only if doing so is feasible for them and does not disrupt their workflow.
3. Filter ratings by what the tool produced and for whom, not just a single overall score. A feature might work well for P4 maths and be quietly failing P2 literacy; an average across all users would hide that difference entirely.
4. Visualise responses to a dashboard that your academic team and the product team both read so that the time a teacher spent rating the tool contributes to your design thinking.
5. Read the open comments. Quantitative insights may point to an issue; but qualitative comments are where you find out what it is, and how it might be fixed.
6. Watch the ratings and comments on new features specifically. A feature that rates below the product average soon after launch is worth investigating.

WATCH OUT

Ratings are a smoke alarm, not a diagnosis. A three out of five tells you where to look, not what to change. In-tool signals often require a return to teachers in-person for deeper understanding. Watch too for prompt fatigue and consider sampling (e.g. asking after every third or fifth use).

CASE STUDY FROM THE SANDBOXES

TALEEMABAD, PAKISTAN

Lesson Genie automatically triggered an in-bot survey after each lesson plan was generated, collecting both a quantitative rating and open qualitative responses, tracked through a dedicated feedback dashboard reviewed by the design team including academic staff and product developers. Teachers rated it 3.8 out of 5 overall, and 4.2 out of 5 specifically for the multi-grade features, the part they had shaped most directly (↑Burgess, 2026).

Validate

10. Usage data review with teachers

Take the pattern you see in the backend data back to the teachers who created it, and ask them what it means. The data tells you what; only teachers tell you why.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
Whenever usage data surprises you	**Developer brings data, teachers interpret it**	**1 to 2 weeks, including school visits**	**5 to 15 teachers**

HERE'S HOW

1. Design the backend of your tool so that you can see which features are used, by whom, how often, and whether people come back.
2. Look for the patterns you did not predict, or check metrics set for specific indicators. Monitor data over time to consider whether patterns represent a trend or do not act on the data alone. Write down your best guess at the explanation, then go and find out whether or not you are right.
3. Go back to schools and show teachers the patterns. “We can see people are doing this — what is going on?” Teachers will provide insights that would be challenging to identify from the data alone.
4. Feed the explanations you found directly into the next sprint download. Consider what design or implementation choices might allow you to respond directly to the challenges identified.

WATCH OUT

Usage data is the easiest thing in this toolkit to over-read, because it arrives clean, quantified, and without arguing back. It is also the method where extraction risk is highest: teachers generate this data continuously and by default. Does a teacher know their in-tool usage is being reviewed?

CASE STUDY FROM THE SANDBOXES

RISING ACADEMIES, RWANDA AND TALEEMABAD, PAKISTAN

Rising Academies' backend data showed teachers switching from English to Kinyarwanda when talking to Tari — unexpected, because Rwanda mandates English as the medium of instruction. Rather than act on the data alone, the academic team went back to schools and found teachers were privately using Tari to rehearse English pronunciation before students arrived. They built a translation feature; today 30% to 40% of Tari conversations include a translation request, while English remains the medium of instruction. Taleemabad tracked usage continuously via PostHog, giving the development team real-time visibility into which features teachers used and whether they returned (↑Burgess, 2026).

Validate

11. Teacher experiment cycles

Teachers are provided the tool (or feature) to experiment directly with in classroom settings. For instance, teachers can be asked to plan, deliver, and observe a lesson together, with and without an AI tool to identify the gap between how you think the tool is used and how it actually is.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
When you need classroom evidence, not self-report	**Teachers, in an existing professional structure**	**4 to 8 weeks, 3 to 4 cycles**	**4 to 8 teachers per cycle**

HERE'S HOW

1. Anchor the experimentation cycle in a professional structure that already exists in the system, rather than creating a parallel one. Many education systems have pre-existing communities of learning, cluster meetings, or subject groups, that already have time set aside for collaborative learning and reflection.
2. Have the group take on a task together using the AI tool as they would normally, and record how they engaged with it (what they submitted as prompts) and the relevant output.
3. Reflect on the experience of leveraging an AI feature or tool for the given task. Compare against the traditional way that teachers may have performed the given tasks. Where did the tool's features or outputs hold, where did it break, and did the AI-generated element help or get in the way?
4. Have developers consider these cycles of experimentation as they develop new features or consider adjustments to the tool's design and implementation. They are the quickest tool for identifying the gap between how you designed the tool to be used and how it is actually used.

WATCH OUT

Frequent experimentation with new features or tools may lead to important iterations to your product, but consider the disruptions classroom practice or to continuous professional development sessions and gauge carefully how often and when to run these cycles.

CASE STUDY

EVIDENCE FROM KENYA

Our colleagues found a persistent gap in Kenya between how designers theorised teachers would use a digitally personalised learning tool and what teachers actually did with it in classrooms: teachers consistently repurposed the tool according to their own pedagogical logic. The finding parallels all three sandboxes, and it is the argument for structured classroom observation over self-report. It also reinforces the wider point that sustained engagement across a design cycle, rather than a one-off consultation, is what generates actionable insight (↑Daltry et al., 2025).

Champion

12. Teacher as product champion

Teachers who shaped a tool can also be the ones who spread it. Adoption through peer networks moves faster, and sticks harder than a top-down rollout.

WHEN TO USE IT	WHO LEADS	TIME REQUIRED	GROUP SIZE
After teachers have visibly shaped the product	**Teachers, entirely**	**Weeks to months, largely self-sustaining**	**Starts with 10 to 20, spreads outward**

HERE'S HOW

1. Recognise that this method cannot be started here. It is the pay-off from the earlier stages. Teachers are more likely to champion tools they helped build.
2. Make each teacher's influence on the product visible and attributable, in public communities and design workshops. "I asked for this and they built it" is the sentence that travels.
3. Cultivate champion users through networks teachers already have, rather than through your distribution list. The relationships that carry adoption are the ones that already exist.
4. Remove friction for the person doing the sharing. Track where uptake comes from, so you can see whether growth is organic or driven by your own communications.

WATCH OUT

Champions are not free labour. If teachers are effectively running your adoption strategy, be honest about that and recognise it properly. And watch for the champion trap: if one enthusiastic teacher becomes the only person who understands the tool in their school, you have created a dependency, not an adoption.

CASE STUDY FROM THE SANDBOXES

TALEEMABAD, PAKISTAN AND AGA KHAN FOUNDATION, TANZANIA

Lesson Genie scaled from 17 teachers to 200 organically, through peer-to-peer sharing, without top-down pressure from Taleemabad's central team — faster than a previous, more top-down rollout by the same organisation. Teachers were also seen using the tool unprompted during the winter school holidays.

In Tanzania, lead teachers were colleagues facing the same constraints as their peers; when one said "I used this, and it worked", word spread. Teachers from Arusha and Morogoro — regions entirely outside the sandbox scope — joined experiments through informal networks with no formal invitation, and materials prepared by seven teachers in one district were accessed by teachers in two further districts (↑Burgess, 2026).

Beyond co-design with teachers

We believe that teacher-led design can help produce tools that work in classrooms, and better test implementation of those tools to meet the needs of teachers. We hope that by providing guidance on how to leverage existing evidence and human-centered design methods towards AI tool development that is more responsive to teachers' unique needs and expertise, this toolkit contributes to AI uses in education that are more effective, equitable, and sustainable.

However, teacher-centered tools cannot, by themselves, shift the ecosystems in which these tools are applied, including addressing policy and infrastructure limitations. These are not teacher-level problems. They require ministry-level action on policy, curriculum requirements, and language model development. Working towards more equitable and effective education in the age of AI requires working towards these challenges simultaneously.

LEARNING OUT LOUD

Building AI tools with teachers, rather than for them, is happening in parallel across many countries, but the learning mostly stays local. A ministry piloting an AI tool at district scale, a developer iterating with teachers each sprint, an implementing partner running co-design in schools — each is generating evidence that the next team doing similar work could use, but has no easy way to find.

We think the field needs faster, more direct ways for teams to compare notes while projects are still underway, rather than waiting for a report after the fact. This toolkit is offered in that spirit: a starting set of methods, not a finished one. Tell us what you tried, what worked, and what didn't — we'll use it to keep the toolkit current and useful to the next team doing this work.

GET IN TOUCH

Tell us what you are building, what you have tried, and what you have learnt. And let us know if you would like to be part of a collaborative on teacher-led AI design. Write to the AI Observatory at **aiobservatory@edtechhub.org.**

References

Adam, T. (2026). *How Is AI Transforming Teachers' Roles in Low- and Middle-Income Countries?* EdTech Hub. https://doi.org/10.53832/edtechhub.1152 Available from https://docs.edtechhub.org/lib/95DU3BRT Available under Creative Commons Attribution 4.0 International.

Adam, T., & Lester, J. (2025). *How Is AI Upgrading Teachers' Roles in Low- and Middle-Income Countries?* EdTech Hub. https://doi.org/10.53832/edtechhub.1150 Available from https://docs.edtechhub.org/lib/36N3ZRFD Available under Creative Commons Attribution 4.0 International.

Adam, T., & Lester, J. (2026). *How Is AI Disrupting Teachers' Roles in Low- and Middle-Income Countries?* https://doi.org/10.53832/edtechhub.1151 Available from https://docs.edtechhub.org/lib/UXT3K2K3 Available under Creative Commons Attribution 4.0 International.

Anaissie, T., Cary, V., Clifford, D., Malarkey, T., & Wise, S. (2026). *Liberatory Design Framework & Toolkit*. Stanford d.School. https://dschool.stanford.edu/tools/liberatory-design

Bastani, H., Bastani, O., Sungu, A., Kabakcı, Ö., & Mariman, R. (2024). *Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics* (SSRN Scholarly Paper No. 4895486). Social Science Research Network. https://doi.org/10.1073/pnas.2422633122

Burgess, F. (2026). *Teachers-in-the-Lead: Insights from Three Sandboxes*. EdTech Hub. https://doi.org/10.53832/edtechhub.1234

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press. https://mitpress.mit.edu/9780262043458/design-justice/

Daltry, R., Hinks, J., Sun, C., Major, L., Otieno, M., & Otieno, K. (2025). Integrating Digital Personalised Learning into Early-Grade Classroom Practice: A Teacher–Researcher Design-Based Research Partnership

in Kenya. *Education Sciences*, *15*(6), 698. https://doi.org/10.3390/educsci15060698

EdTech Hub. (2022). *Sandbox Handbook v2.0*. EdTech Hub. https://doi.org/10.53832/edtechhub.0108

Google Ventures. (2019). *The Design Sprint—GV*. http://www.gv.com/sprint

Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). *Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI- Math Tutor in Ghana*. https://arxiv.org/pdf/2402.09809

Holstein, K., & Aleven, V. (2022). Designing for human–AI complementarity in K-12 education. *AI Magazine*, *43*(2), 239–248. https://doi.org/10.1002/aaai.12058

IDEO. (2015). *The Field Guide to Human-Centered Design*. IDEO.org. https://design-kit-production.s3.us-west-1.amazonaws.com/Field_Guides/Field+Guide+to+Human-Centered+Design_IDEOorg_English.pdf

Roy, P., Poet, H., Staunton, R., Aston, K., & Thomas, D. (2024). *ChatGPT in Lesson Preparation: A Teacher Choices Trial*. Education Endowment Foundation. https://www.isrctn.com/ISRCTN13420346

Sheffey, A. (2025). *South Korea's teachers revolted against AI. It's a cautionary tale for the US*. Business Insider. https://www.businessinsider.com/ai-in-school-south-korea-textbook-rollback-jobs-education-2025-8

Sungu, A., Lira, B., & Duckworth, A. (2026). *Generative AI Can Harm Teaching* (SSRN Scholarly Paper No. 7007339). Social Science Research Network. https://papers.ssrn.com/abstract=7007339

Teacher Task Force. (2025). *Promoting and Protecting Teacher Agency in the Age of Artificial Intelligence* (The International Task Force on Teachers for Education 2030, p. 36). UNESCO. https://teachertaskforce.org/sites/default/files/2025-09/1149_25_Promoting%20and%20Protecting%20Teacher%20Agency_FINAL_3Sep.pdf

USAID, & Aga Khan Foundation. (2025). *Accelerate impact guide to*

human-centred design for social innovation. Aga Khan Foundation Learning Hub. https://akflearninghub.org/initiative/human-centred-design/all-phases/