

# Request for proposals

## Evaluation guide for testing EdTech/AI tools

October 2026

**Opening date:** 23<sup>rd</sup> October 2026

**Closing date for proposals:** 13<sup>th</sup> November 2026

---

### Background and context

The Department for Education (DfE) and the Education Endowment Foundation (EEF) have partnered to deliver the [Education Technology \(EdTech\) Testbed Programme](#) (2026-2030), a new multi-year research initiative dedicated to evaluating emerging EdTech tools with the potential to improve teaching, learning and inclusion outcomes in education.

EdTech tools are digital technologies harnessed into programmes and approaches to support teaching and learning, including computer adaptive learning and artificial intelligence (AI) such as large language models (LLMs). These tools offer opportunities to enhance teaching and learning experiences, widen access to learners, improve efficiencies and deliver value for money for schools and colleges. As an increasing variety of EdTech products enter the market, there is a growing need for timely and actionable evidence to inform decisions about investment, adoption and scaling of EdTech in classrooms.

The EdTech Testbed Programme sets out to support this aim by strengthening the evidence base on the impact of EdTech in the following three key priority areas:

- Improving learner outcomes
- Supporting teaching and administrative workload
- Enabling inclusion of all learners, including those with special educational needs (SEN)

The EdTech Testbed Programme comprises two complementary stages. Stage 1 will be delivered by a research consortium led by Ipsos, which will conduct up to 75 rapid evaluations aligned with the Testbed Programme's priority areas. These studies will focus on tools that are feasible for testbed evaluation and eventual progression to larger-scale Stage 2 evaluations.

Between October 2026 and April 2027, approximately 15 tools will be tested as part of the first cycle of Stage 1 rapid evaluations, with further cycles in Stage 1 spanning the duration of the Testbed Programme and promising tools progressing into Stage 2.

From spring 2027, the [EEF will lead Stage 2 of the Testbed Programme](#), conducting up to 15 large-scale robust evaluations of the tools and programmes from Stage 1 demonstrating the most promise.

To date, the EEF has commissioned a range of efficacy and effectiveness experimental trials as part of its expanding [EdTech](#) research theme. Recently published research [includes two Teacher Choice trials](#) investigating the impact of teachers using AI lesson planning tools on teacher workload as well as three programme evaluations of EdTech: [Maths Whizz](#), [Lexia](#) and [Reading Plus](#). Insights from these recently completed studies and wider emerging research present a

timely opportunity for synthesising lessons and arising methodological challenges into a distilled set of key considerations to effectively design and run EdTech evaluations.

The diversity of EdTech products, use cases and evidence questions that will emerge from Stage 1 of the Testbed Programme also necessitate a coherent set of evaluation considerations to support evaluators in making effective, relevant and proportionate design decisions. A central challenge for Stage 2 of the EdTech Testbed programme will be to balance EEF's standards for methodological rigour with the need to produce timely and relevant evidence.

## Purpose and scope of work

The EEF is seeking to appoint a research team to develop a practical, evidence-informed guide to support EEF evaluations of EdTech/AI tools.

Stage 2 evaluations are expected to span all education key stages and subjects, including primary and secondary schools, as well as 16-19 educational settings. A combination of efficacy and effectiveness trials will be commissioned to test the most promising tools from Stage 1 in the following two key areas of the Testbed Programme:

- **Teacher and administrative workload** - testing tools designed to support practitioners with planning, instructional and administrative practices.
- **Pupil attainment** - testing tools that enable adaptive instruction and personalised tutoring for learners of all levels of ability. This will include developing an understanding of attainment outcomes for pupils with SEN and pupils eligible for free school meals (FSM).

The main deliverable from this project will be a concise, practical guide to support evaluations of EdTech/AI tools. Given the breadth of use cases covered across the Programme's priority areas, we expect that EdTech tools progressing into Stage 2 will vary in product maturity, implementation contexts, intended outcomes, and the strength of prior product, implementation and process evaluation insights generated through Stage 1. EdTech and AI tools that progress to Stage 2 will satisfy standard product and evaluation readiness criteria, ensuring that their delivery models are sufficiently codified and developed for more rigorous evaluation. These standards are still under development.

A successful guidance output will support evaluators to navigate a range of study contexts and make appropriate decisions regarding trial design, implementation monitoring, outcome selection and measurement, as well as effective engagement with EdTech developers. The guide should provide a coherent set of considerations for evaluation design decisions, whilst maintaining flexibility to accommodate different product, setting and evaluation contexts. We expect the guide to focus on the following design and methodological considerations:

- A. Selecting an appropriate evaluation design
- B. Selecting and measuring outcomes
- C. Defining and examining implementation and use pathways
- D. Working effectively with EdTech developers

## **A. Selecting an appropriate evaluation design:**

- Selecting RCT designs appropriate to the study context and aligned with EEF standards, including when multi-arm, sequential, crossover, adaptive and A/B trial configurations may be better suited.
- Choosing the most appropriate inferential framework, such as superiority, equivalence, or non-inferiority testing.
- Defining [core components](#) and causal mechanisms underlying the key tasks and processes being evaluated, informed by prior theory validation undertaken during Stage 1 and building on existing EEF guidance relating to core components and [theories of change](#). This should include capturing evidence on EdTech-specific core components such as personalised resource generation, adaptive and scaffolded learning content, self-paced learning, individualised feedback, and workload reallocation, as well as how they relate to the non-technological planning and instructional core components underlying the interventions.
- Identifying appropriate target populations of schools and teachers and defining plausible comparison conditions that support experimental contrast and feasible implementation adjustments for schools. Comparison conditions may range from pure business-as-usual conditions, non-EdTech/AI time-equivalent alternatives, to selective EdTech/AI usage conditions.
- Determining the degree of implementation control, informed by Stage 1 evidence regarding expected levels of adoption, engagement and usage.
- Setting the most appropriate duration for the trial with emphasis on shorter 6–12-week implementation timelines for proximal outcome testing, with options for longer-term collection of more consequential outcomes.
- Considering how the design considerations above should be calibrated based on whether an efficacy or effectiveness level evaluation is being conducted.

## **B. Selecting and measuring outcomes**

- Identifying well-suited proximal and longer-term, distal outcome constructs and choosing valid measurement approaches that balance practical burden and measurement quality. Assessment approaches may include rapid or repeated testing of proximal outcomes over shorter timelines. Outcomes will differ depending on whether the evaluation is focused primarily on teachers or pupils. Typical proximal outcomes may relate to immediate and near-transfer task performance measured through diagnostic assessments of knowledge and accuracy, as well as short-term shifts in workload for teachers measured through workload diaries and usage analytics.
- Exploring heterogeneous impacts for different school and teacher subgroups as well as those based on other moderating and mediating contextual factors.

### **C. Defining and examining implementation and use pathways**

- Integrating platform-generated data into analysis to examine frequency, depth and patterns of engagement. This should include recognising the limitations of platform data as a measure of ultimate outcomes and avoiding over-reliance on metrics tied solely to intervention conditions.
- Establishing expectations of adoption, acceptability, compliance, implementation fidelity, and adaptations, as well as the means to measure and contextualise these aspects of implementation with downstream outcomes.
- Utilising a diversity of sources such as platform data, observation, self-report and assessment data to validate specific outcomes or connected segments of the theory of change.

### **D. Working effectively with EdTech developers**

- Managing communication and decision-making where developers evolve or modify tools during an evaluation. This includes defining acceptable product adaptations and setting expectations for which product changes are permissible during the evaluation, documenting modifications and factoring these into the assessment of intervention fidelity and outcomes.
- Accessing product and implementation data from developers to inform design assumptions and benchmarks for evaluation of tools.
- Working with developers to articulate theories of change and underlying task behaviours to be evaluated.
- Establishing data sharing, governance and data security arrangements to support evaluation.

## **Main deliverable**

The main deliverable from this project will be a concise, practical guide for evaluations of EdTech/AI tools. We expect the resulting guide to satisfy the following expectations:

- The guide should be structured around the four topic areas presented in the previous section of this tender. Applicants may propose additional topic areas where appropriate. Any proposal to omit one or more of the four topic areas should be well-justified.
- Key considerations within each topic area should be informed by lessons learnt from EEF-commissioned and other EdTech/AI evaluations, providing illustrative project examples where appropriate. We recognise that the guide may include some reflection on emerging and unprecedented methodological challenges and opportunities, particularly in relation to evaluating AI tools.
- We recommend the guide feature an outcome measures framework segmented for each priority area of the Testbed Programme (learner outcomes, teacher workload and

inclusivity). Within each priority area, the framework should identify the most common and influential conceptual mechanisms, immediate proximal and longer-term distal outcomes, and viable measurement approaches, highlighting the strengths and limitations of different measurement approaches.

- The production of any supplementary tools, frameworks, checklists or supporting resources should be appropriately justified.
- Whilst primarily intended for evaluators, the guide should be accessible to EEF, DfE and other stakeholders involved in commissioning, reviewing and interpreting Stage 2 evaluations.
- The guide should align with and build on existing methodological and process related resources produced by both the EEF and the DfE (more in next section).

## Methodological expectations

Interested applicants should set out their proposed methodology for producing the guide for the evaluation of EdTech/AI tools. We expect that a desk-based review and synthesis of the following resources will form the core part of the preparatory work to develop the evaluation guide:

- EEF internal EdTech evaluation and methodological guidance, [EdTech evidence reviews](#), individual evaluation reports as well as more general [EEF methodological and process resources](#) relating to core components, [theories of change](#), [statistical analysis](#), and [scale up](#).
- Relevant DfE [testbed evaluation guidance](#), as well as overarching and project-level resources (study protocols and pre-pilot findings) from the first cycle of Stage 1 rapid evaluations between October and December 2026. Relevant resources and updates from Stage 1 of the Testbed programme will be shared with the appointed team at project set-up and at appropriate intervals across the project.
- Other EdTech framework and standards, including the [Magenta Book](#), EdTech/AI evaluations and available literature.

Applicants may suggest other research activities but should propose an overall proportionate approach that supports the timely production of the stated deliverables. Project management activities will include an initial set-up meeting, regular progress meetings with the EEF, and revisions to the main deliverable in response to EEF and peer review.

## Indicative timeline

### Call for proposals (October – November 2026)

- Application window opens (23<sup>rd</sup> October)
- Application closes (13<sup>th</sup> November)
- Application scoring and appointment (20<sup>th</sup> November)

### Appointment and project set-up (November 2026)

- Project set-up meeting (final agreement on scope of work, deliverable, timelines and budget)
- Due-diligence and grant agreement

### Development of evaluation guide (December 2026 – January 2027)

- Desk-based research and guidance development
- Routine project meetings with EEF

### Submission and review of draft deliverable (February 2027)

- Submission of draft main deliverable output
- EEF and peer review
- Proofreading and design

### Submission of final deliverable (March 2027)

## Budget

We anticipate an indicative project budget in the range of £15,000-£25,000 for work undertaken by a single applicant researcher or small team. Proposals will be assessed on their value for money and proportionality in relation to the proposed methodology for developing the evaluation guide. While we expect the project to fall outside the scope for VAT, the responsibility for the VAT assessment sits with the provider.

## Commissioning process

Proposals will be scored holistically across the following criteria: **Capability and relevant experience of the core project team; research methodology and approach; approach to research governance**, and **value for money**.

| Scoring Criteria |                                                                     |
|------------------|---------------------------------------------------------------------|
| 0                | Totally fails to meet the requirement - information not available   |
| 1                | Meets some of the requirements with limited supporting information  |
| 2                | Meets some of the requirements with reasonable explanation          |
| 3                | Fully meets the requirements with detailed explanation and evidence |
| 4                | Exceeds the requirements with extensive explanation and evidence    |

- **Capability and relevant experience of core project team:** The applicant has appropriate expertise in EdTech/AI research, impact, implementation and process evaluations, prior experience in developing methodological guidance as well as project management capabilities.
- **Research methodology and approach:** The proposal provides a clear, proportionate and well-justified approach to developing the evaluation guide, drawing on relevant evidence

with a feasible plan for producing a practical resource tailored to the end-user audience with clearly described intended use.

- **Approach to research governance:** The proposal provides appropriate risk mitigations and quality assurance measures to support alignment to project aims and requirements.
- **Value for money:** The proposal includes a clear and proportionate budget demonstrating efficient use of resources and value for money in delivering a high-quality output within required timeframes.

### Commissioning and project timeline

|                                                     |                                                   |
|-----------------------------------------------------|---------------------------------------------------|
| <b>Deadline for submitting evaluation proposals</b> | 5 PM, 13 <sup>th</sup> November 2026              |
| <b>Appointment of evaluator</b>                     | 20 <sup>th</sup> November 2026                    |
| <b>Set-up meeting</b>                               | 23 <sup>rd</sup> – 27 <sup>th</sup> November 2026 |
| <b>Guidance development</b>                         | December – January 2027                           |
| <b>Draft deliverable</b>                            | February 2027                                     |
| <b>Final deliverable</b>                            | March 2027                                        |

### Submissions

Interested applicants should prepare and [submit a proposal](#) up to 2,500 words (including all tables and appendices, but not budget or references) by **5pm on 13<sup>th</sup> November 2026**.

If you have any questions, please contact Faizaan Sami [faizaan.sami@eefoundation.org.uk](mailto:faizaan.sami@eefoundation.org.uk)