

Technical Report

A randomised controlled trial investigating the impact of Eedi versus typical teaching approaches

Mathematics attainment in KS3
(Year 7)

I= Eedi

Full Technical Report 2026

Dr Emma Dobson, Rahil Khowaja, Dr Wayne Harrison, Germaine Uwimpuhwe,
Professor Steve Higgins

Contents

Executive Summary	2
Background	4
Methods.....	11
Impact Results: Summary.....	16
Impact Results: Technical Analysis.....	17
IPE Survey Results: Summary.....	24
IPE Survey Results: Technical Analysis.....	26
IPE Interview Results	49
Discussion	62
Conclusion	66
References.....	67

Executive Summary

This evaluation presents the findings from a randomised controlled trial (RCT) evaluating the impact of Eedi, an online diagnostic mathematics platform, on Year 7 attainment. The study examined both pupil outcomes and implementation processes across 16 secondary schools. A total of 2,296 pupils participated in the study, with classes randomly assigned within schools to either intervention groups using Eedi for homework and/or classwork, or control groups continuing with business-as-usual mathematics teaching.

Key Findings

- **Positive overall impact:** Pupils using Eedi scored slightly higher on the post-test than those receiving business-as-usual teaching ($g = 0.10$, 95% CI -0.08 to 0.28). This difference was not statistically significant ($p = 0.27$): a small positive signal under real-world classroom conditions.
- **Stronger impact with sustained use:** Estimated effects rose with the level of engagement, a pattern consistent with a dose-response relationship. Pupils who engaged more frequently with Eedi achieved larger attainment gains.
- **Substantial gains for highly engaged pupils:** At the highest levels of engagement (200+ diagnostic questions), the effect increased to $g \approx 0.24$ (95% CI 0.03 to 0.44), reaching statistical significance. These compliance analyses are exploratory.
- **Particularly strong results for disadvantaged pupils:** Among highly engaged Pupil Premium pupils the estimated effect was substantial ($g \approx 0.57$, 95% CI 0.32 to 0.82). This is an exploratory complier estimate based on the 36% of Pupil Premium pupils who reached the highest engagement threshold; it is promising rather than proven, but points to potential to support closing the attainment gap when implementation is strong.
- **Teacher endorsement:** Most teachers found Eedi easy to use and reported no detrimental impact or a positive reduction in workload.

This evaluation found a positive but not statistically significant average effect of Eedi on mathematics attainment under typical classroom conditions, alongside converging exploratory evidence that estimated effects strengthen with sustained pupil engagement. In particular, the substantial estimated gains among highly engaged disadvantaged pupils point to Eedi's potential to contribute to narrowing attainment gaps. The results suggest that when embedded consistently within curriculum

planning and supported by strong implementation, Eedi may offer a scalable, evidence-informed approach to enhancing learning outcomes in secondary mathematics.

Background

Context

Improving attainment in Key Stage 3 mathematics remains a persistent challenge, particularly as gaps in foundational understanding can compound over time. In this context, educational technology (EdTech) is increasingly positioned as a scalable way to support learning through individualised practice, rapid feedback, and diagnostic-informed teaching. The evidence base suggests EdTech can improve attainment, but typically by a modest margin and under specific conditions.

Two recent syntheses illustrate the overall pattern: EdTech interventions tend to deliver small average gains in mathematics (around $g \approx 0.2$) (Di Pietro & Castaño Muñoz, 2025; Ran, Kim, & Secada, 2022). However, effect sizes vary considerably depending on how technology is used. Larger gains are more likely when technology supports higher-level learning processes (e.g., conceptual understanding, problem-solving, collaborative work) than when it is used mainly for repetitive drills (Hillmayr et al., 2020). This aligns with wider evidence showing that technology added to classrooms without strong integration into teaching and learning routines often produces limited or no measurable benefit (J-PAL Evidence Review, 2019). The implication is that the impact of EdTech is implementation-dependent: benefits are most likely when platforms are used in ways that complement pedagogy rather than functioning as a bolt-on replacement.

Within the broad category of EdTech, systems that combine diagnostic assessment, adaptive practice, and feedback are especially relevant to mathematics. Evidence from multiple contexts suggests that computer-assisted and intelligent systems can improve outcomes when they support learners at the right level and help teachers respond to misconceptions. Personalised mathematics software frequently improves attainment, though results are mixed. The J-PAL review of education technology evaluations reported that around two-thirds of rigorous studies of computer-assisted learning showed significant positive effects (J-PAL Evidence Review, 2019). One well-cited example is the ASSISTments platform: in a randomised trial, students using ASSISTments for online homework made significant gains after one year (≈ 0.18 SD),

with stronger benefits for lower-attaining pupils; consistent with the idea that feedback and scaffolding matter most when learners are struggling (Roschelle et al., 2016).

Mathematics learning is particularly vulnerable to the accumulation of misconceptions: small misunderstandings can cascade into broader failure to access later content. Longstanding work in mathematics education shows that misconceptions, if left unaddressed, can obstruct progress and reduce learners' confidence and persistence (Durkin & Rittle-Johnson, 2015). The broader evidence indicates that diagnostic information can improve learning. For example, providing teachers with diagnostic feedback on students' mathematical skills has been shown to improve subsequent attainment (≈ 0.1 – 0.2 SD), particularly when diagnostics are repeated and used to guide ongoing instructional decisions (Betts, Hahn, & Zau, 2017).

A second mechanism that appears critical in successful EdTech is timely feedback. Immediate explanations, prompts, or hints reduce the risk that pupils rehearse errors and can support persistence through challenging material. Meta-analytic evidence indicates stronger impacts when technology functions as an instructional partner, adapting to student responses and providing feedback, rather than simply delivering content (Ran et al., 2022). Many modern platforms achieve this adaptivity through artificial intelligence, including models of student knowledge and algorithms that recommend what a learner should do next. Evidence on intelligent tutoring systems suggests they can outperform business-as-usual instruction and, in some cases, approach the impact of human tutoring (Ma et al., 2014). At the same time, the emerging AI-in-education literature highlights challenges that are especially relevant to school systems including the risk of widening inequalities when infrastructure and access are uneven (Meylani, 2025; J-PAL Evidence Review, 2019). These concerns have sharpened interest in "human-in-the-loop" models, where AI supports learning and diagnosis, but teachers remain responsible for interpretation, integration into curriculum, and equitable implementation (Luckin et al., 2016). Early evidence from a recent evaluation of Eedi suggests that this blend may be educationally meaningful; in a two-year trial, Eedi's average effects were positive across timepoints and became significant at longer follow-up, alongside a strong compliance (dose–response) pattern (Harrison et al., 2025). This hints at a plausible hypothesis consistent with the wider literature: impacts may be modest early on, but become clearer when use is sustained and consistent.

Gaps in the existing evidence base

Despite a generally positive direction of findings, several gaps limit what is currently known about diagnostic and AI-enabled EdTech in real school contexts. Some meta-

analyses report larger impacts of blended or online mathematics instruction (e.g., Akin, 2022; Samritin et al., 2023; Setiawan et al., 2022; Putri et al., 2025). However, these effects frequently come from short interventions, small samples, intensive researcher support, and non-standardised assessments. These conditions can inflate effect sizes and reduce generalisability (Means et al., 2013; Escueta et al., 2020; Hillmayr et al., 2020; Pellegrini et al., 2021). In more ecologically valid settings, effects are typically smaller and more variable (Benavides-Varela et al., 2020; Güler et al., 2023). This creates uncertainty about what schools should expect from commercial platforms implemented under everyday constraints.

A consistent limitation in EdTech impact trials is that the “active ingredient” is underspecified: many studies report an intent-to-treat estimate without robust reporting on how often tasks were set, how consistently pupils engaged, and how closely use aligned with curriculum goals (Kabudi, Pappas, & Olsen, 2021; Létourneau et al., 2025). Yet fidelity plausibly determines impact, particularly for platforms that rely on cumulative retrieval practice and feedback loops. Without detailed dosage and implementation measures, null or small effects are hard to interpret (O’Donnell, 2008), and variability between schools remains unexplained. Existing work suggests teacher engagement and instructional alignment can substantially shape uptake and outcomes (Chen et al., 2020), but too few peer-reviewed evaluations combine achievement outcomes with rich platform analytics and explanatory qualitative data.

The evidence on whether EdTech reduces or widens attainment gaps is still limited. Some syntheses suggest disadvantaged learners can benefit at least as much as their peers, and possibly more in maths and science (Di Pietro & Castaño Muñoz, 2025). However, digital inequality remains a major constraint: disparities in device access, home connectivity, and study environments are well documented, and disadvantaged pupils often face more barriers to engaging with online homework and remote learning (Andrew et al., 2020; Maldonado & De Witte, 2022; Feng et al., 2023). International reviews warn that when access is uneven or when platforms assume substantial out-of-school use, EdTech can unintentionally widen gaps (Escueta et al., 2020; UNICEF, 2025). Many impact evaluations give limited attention to these equity-relevant mechanisms (Galimova et al., 2024; Feng et al., 2025), leaving a gap in understanding *for whom* diagnostic EdTech works and under what conditions.

Finally, there remains a shortage of independent, peer-reviewed evaluations of commercial diagnostic platforms that include (a) standardised attainment outcomes, (b) real-world implementation, and (c) long-term follow-up. Reviews highlight that relatively few studies combine these elements in a single design (Scherer et al., 2019;

van der Kleij, 2019). This constrains the field's ability to draw conclusions about sustained impact, scalability, and the implementation support required for success.

Eedi's Operational Theory of Change

Eedi is an innovative EdTech platform offering over 60,000 diagnostic maths questions, engaging 15,500 monthly users through its school and tutoring services. Enhanced by AI and video explanations, Eedi helps learners tackle misconceptions and provides additional support. Premium users benefit from live, one-on-one tutoring, allowing students to request and receive assistance through the platform's chat function. Eedi's operational theory of change brings together its platform assets, support features, and school context to outline how improved teaching and learning outcomes are expected to occur.

Inputs and Context

Eedi provides teachers with a large bank of curriculum-aligned diagnostic quizzes, misconception analytics, and pupil engagement features (e.g., leaderboards and rewards). Pupils are supported through scaffolded hints, explainer videos, fluency practice and access to hybrid human tutors. Successful implementation depends on contextual factors including device and internet access, teacher autonomy, and school leadership support.

Teacher Adoption

Teachers must reach a routine of at least 20 active weeks per year. Adoption follows the Technology Acceptance Model:

- Ease of Use (PEOU): Automated marking lowers workload and reduces barriers to adoption.
- Perceived Usefulness (PU): Teachers gain access to misconception data unavailable through traditional methods.
- Adaptation: Teachers use these insights to adjust instruction, reinforcing a habit loop of continued use.

Student Dosage

Impact requires students to complete approximately 120 diagnostic questions per year. Four mechanisms underpin this:

- Shared Reality: All pupils complete the same diagnostic quiz.

- Low-Stakes Formative Assessment: Quizzes embed spacing, retrieval and active learning principles.
- Adaptive Remediation: Personalised lesson experiences target specific misconceptions, supported by optional hybrid tutoring.
- Retrieval Practice: Personalised retrieval quizzes reinforce previously learned content.

Expected Outcomes

Proximal outcomes include improved teacher insight into misconceptions, more targeted differentiation, and more efficient planning. Intermediate outcomes include increased student mastery, improved self-efficacy, and more effective data-driven instruction. Distal impacts include improved mathematics attainment, reduced FSM/non-FSM attainment gaps, reduced teacher burnout, and strengthened teacher professional efficacy.

Theoretical Basis

Eedi is underpinned by a clearly articulated theory of change that integrates three well-established strands of educational research: cognitive science, implementation science, and equity-focused learning theory. Together, these components provide a coherent rationale for how the platform is intended to improve mathematics attainment and support effective classroom practice.

Cognitive Science: Intended Learning Mechanisms

Eedi's design assumes that formative diagnosis should precede practice. Two mechanisms from cognitive science are central. First, retrieval practice: the regular use of low-stakes diagnostic quizzes is intended to activate the testing effect, a process shown to support long-term retention more effectively than restudy (Roediger & Karpicke, 2006). Second, misconception resolution: drawing on cognitive load theory, the platform aims to identify specific misconceptions and provide scaffolded, corrective feedback to address them (Sweller, 1988). This is intended to prevent pupils from practising errors and to reduce intrinsic cognitive load.

Implementation Science: Intended Adoption Pathways

Eedi's theory of change recognises that impact depends on reaching a sufficient level of use. Prior implementation research suggests that EdTech effects are often non-linear, with a minimum "dosage" required before benefits appear (Fixsen et al., 2005). Eedi operationalises this through weekly curriculum-aligned quizzes, supported by

previous RCT evidence indicating that approximately 120 diagnostic questions per year are associated with improved attainment (Harrison et al., 2025). Teacher adoption is framed through the Technology Acceptance Model (TAM): ease of use is supported through automated marking and analytics; usefulness is expected to arise from access to misconception data that teachers would not otherwise generate.

Equity and Interaction: Intended Support for All Learners

Eedi aims to mitigate widening attainment gaps by promoting differentiation by support, not by task, avoiding the risks of fully adaptive pathways that may limit challenge for lower-attaining pupils (Stanovich, 1986). All pupils complete the same diagnostic quiz, enabling teachers to identify whole-class needs. Personalisation occurs after diagnosis through scaffolded lessons, hybrid tutoring, and targeted practice aligned with the learner’s current understanding (Kucirkova et al., 2025). Chat-based tutoring further seeks to democratise access to high-quality one-to-one support (Nickow et al., 2024).

INPUTS & CONTEXT	TEACHER ADOPTION (TAM)	MECHANISMS	OUTCOMES
<p>Platform Assets</p> <ul style="list-style-type: none"> • Large bank of diagnostic quizzes, teachers assign these to the class in line with their scheme of work • Data-driven instructional support: Misconception analytics dashboard • Engagement hooks: leaderboards, streaks, coins, avatars to encourage dosage compliance <p>Support Assets</p> <ul style="list-style-type: none"> • Hybrid human tutors • Scaffolded hints • Explainer videos • Stretch and challenge practice • Fluency practice <p>Context</p> <ul style="list-style-type: none"> • Device/internet access • Teacher agency • School leadership support 	<p>Gate A: Teacher Routine (Threshold: ≥ 20 active weeks/year)</p> <p>1. Ease of Use (PEOU) Auto-marking removes the "aversive stimulus" of grading admin ↓</p> <p>2. Usefulness (PU) Teacher accesses unique misconception data ↓</p> <p>3. Adaptation Teacher uses data to re-teach; habit loop is formed</p>	<p>Gate B: Student Dosage (Threshold: ≥ 120 DQs/yr)</p> <p>1: Shared reality → All students complete the same diagnostic quiz, aligned to the classroom curriculum</p> <p>2: Low-stakes formative assessment → Grounded in the science of learning (interleaving, spacing, retrieval, active learning)</p> <p>3: Adaptive remediation → When a misconception is diagnosed, Eedi serves a personalised lesson experience to resolve the specific error. → Students can opt in to an interactive dialogue with a hybrid human tutor.</p> <p>4: Retrieval practice → Personalised retrieval quizzes based on students’ past topics and performance</p>	<p>Proximal</p> <ul style="list-style-type: none"> • Teacher: Greater understanding of which students need targeted support • Teacher: Increased differentiation in instruction • Teacher: More efficient lesson planning • Teacher: Improved diagnosis of common misconceptions in class <p><u>Evidenced by:</u></p> <ul style="list-style-type: none"> • Teacher: analytics dashboard usage • Student: Misconception resolution rate <p>Intermediate</p> <ul style="list-style-type: none"> • Student: Increased mastery of topics • Student: Student self-efficacy • Teacher: More effective, data-driven instruction <p>Distal (Impact)</p> <ul style="list-style-type: none"> • Student: Improvement in standardised maths scores • Student: Reduced attainment gap between FSM and non-FSM pupils. • Teacher: reduced burnout from a more manageable workload • Teacher: Increased professional efficacy (Teacher believes they can support low-attainers).

Figure 1. Eedi’s Theory of Change

Research Aim

The aim of this evaluation is to provide a rigorous and comprehensive assessment of the Eedi platform’s effectiveness in improving Year 7 mathematics attainment, while simultaneously enriching the broader evidence base in areas that have been under-researched (long-term effects of sustained use, fidelity, and equity in EdTech). By comparing growth in maths attainment between the Eedi intervention group and the

control group (who may use usual teaching or other maths platforms), the study will rigorously test whether Eedi delivers measurable gains in KS3 maths. In addition, the study incorporates a detailed process evaluation to document how Eedi was implemented in diverse school contexts and to explore the relationship between implementation fidelity and outcomes.

Research Questions

This evaluation is guided by several research questions (RQs) that align with the gaps identified in the literature:

- **RQ1: Impact on Mathematics Attainment.** What is the effect of one year of using Eedi on Year 7 students' mathematics attainment?
- **RQ2: Differential Effect for Disadvantaged Pupils.** Does the impact of Eedi on maths attainment vary for disadvantaged students (specifically, those eligible for Pupil Premium, a proxy for low socio-economic status)?
- **RQ3: Role of Implementation Fidelity (Usage).** Does the effect of Eedi vary according to the level of student engagement or dosage of use (e.g. time spent on the platform)?
- **RQ4: Variation Between Schools and Classes.** How much does the impact of Eedi differ across different schools or classrooms?

In addition to these impact-focused questions, the process evaluation component poses its own questions that delve into implementation and comparison conditions:

- How was the intervention actually implemented in schools?
- What were the enablers and barriers teachers faced in using Eedi?
- What did "business-as-usual" look like in control schools, and did some control schools adopt similar tools (i.e. is there potential contamination or substitute practices)?
- How and why did implementation vary between classrooms, and did this variability affect outcomes?
- To what extent did teachers integrate Eedi content with their regular math lessons?

Methods

Study Design

This evaluation used a randomised controlled trial (RCT) design. Classes taking part in the study were randomly assigned to one of two groups:

1. Intervention group: Year 7 classes that used *Eedi*.
2. Control group: Year 7 classes that continued with their usual maths teaching and did not use *Eedi* during the trial period.

School Recruitment and Randomisation

16 secondary schools agreed to participate in the evaluation. Within each school, Year 7 classes were randomly allocated to either the intervention group, which used *Eedi* for homework and classwork for the remainder of the academic year, or the control group, which continued with business-as-usual practice. Across the 16 schools, the trial involved 2296 pupils, with 1204 in the intervention group and 1092 in the control group, with 74 intervention classes and 76 control classes. The sample included 506 pupils eligible for Pupil Premium (267 intervention; 239 control).

Outcome Measures

The primary outcome was attainment in KS3 mathematics, measured using the Renaissance Star Maths standardised assessment. All Year 7 pupils completed a baseline pre-test and a post-test. Tests are computer-adaptive and benchmarked to US norms to support cross-cohort comparability.

Analysis

The analysis employed multilevel modelling to account for hierarchical data structure (students nested within classes nested within schools) and included intention-to-treat (ITT) and complier average causal effect (CACE) analyses. To examine whether the intervention's effectiveness varied by Pupil Premium (PP) status, an interaction model was fitted that included both the main effects of treatment and PP status, as well as their interaction term. The interaction coefficient tests whether the treatment effect differed significantly between PP and non-PP students. A statistically significant interaction would indicate that the effect of the intervention differs between PP and non-PP students; the direction of any difference depends on the sign of the coefficient.

Complementing the interaction analysis, separate subgroup analyses were conducted for Pupil Premium students and non-Pupil Premium students independently. These models used the same multilevel structure as the primary ITT analysis but were fitted separately within each subgroup. This approach provides direct estimates of treatment effects for each population and their associated confidence intervals to explore how the intervention performs for PP and non-PP students.

To explore context-specific variation in treatment effects, school-level analyses were performed by fitting separate models for each of the 16 participating schools. These analyses examined whether the intervention was more effective in some school contexts than others, providing insight into implementation factors or school characteristics that may moderate treatment effects.

Treatment compliance is measured using the number of diagnostic questions answered, with four distinct compliance thresholds established to examine dose-response relationships. Students are classified as compliant at different levels based on whether they gave a response of at least 60, 90, 120, or 200 diagnostic questions. These thresholds are coded as binary variables and are used in the Complier Average Causal Effect analyses to estimate treatment effects for students who engaged with the intervention at varying intensity levels.

The CACE estimation employed a two-stage instrumental variable approach that uses random treatment assignment as an instrument for actual treatment receipt or compliance. In the first stage, compliance (defined by meeting one of the four duration thresholds) was predicted using treatment assignment along with the covariates included in the outcome model. This first-stage regression estimates the relationship between being assigned to treatment and receiving the intervention at the specified dosage level. The fitted values from this model represent predicted compliance probabilities based solely on the randomised assignment and baseline characteristics, thus purging the compliance measure of any endogenous selection effects that might arise if students who choose to engage more with the intervention differ systematically from those who engage less. In the second stage, post-test scores were predicted using the fitted compliance values from the first stage rather than actual compliance. This model maintained the same multilevel structure as the ITT analysis, with random intercepts for schools and groups, and included the same baseline covariates. The coefficient on predicted compliance in this second-stage model provided the CACE estimate, representing the average treatment effect for students who comply with their assignment at the specified threshold. Because the compliance measure has been instrumented by random assignment, this estimate captures the causal effect of the

intervention for compliers while avoiding the bias that would arise from simply comparing outcomes of students who happened to engage more versus less with the intervention.

All treatment effects are reported both as raw regression coefficients (in the original metric of the outcome variable) and as standardised effect sizes to facilitate interpretation and comparison across studies. The primary standardised effect size measure is Hedges' *g*, which represents the treatment effect in standard deviation units after applying a correction for small sample bias. Hedges' *g* is calculated by first computing Cohen's *d* as the ratio of the treatment coefficient to the total standard deviation, where the total standard deviation is obtained by taking the square root of the sum of all variance components in the multilevel model (see, e.g., Westfall et al. 2014).

Implementation and Process Evaluation

Teacher Surveys

Two online teacher surveys were administered: a mid-point survey in February and an end-point survey in July. The surveys included:

- A mathematics teaching confidence scale, adapted from Hettinger et al. (2022), to capture teachers' self-efficacy and perceived capability in supporting pupil learning.
- Questions on implementation fidelity, including how frequently teachers set Eedi quizzes, whether they used Eedi during lessons, and how they integrated diagnostic information into planning.
- Feature-use items, asking teachers which components of the platform they used (e.g., diagnostic questions, dashboards, retrieval quizzes, in-lesson tools).
- Pupil engagement questions, exploring teachers' perceptions of how pupils responded to Eedi, which groups engaged most consistently, and any patterns of disengagement.
- Barriers and enablers, such as device access, curriculum alignment, competing priorities, or technical issues that influenced ongoing use.
- Workload and experience measures, examining whether teachers felt Eedi changed their marking load, planning time or in-class management.

To ensure confidentiality and compliance with the data-sharing agreement, all surveys were distributed by Eedi directly to participating teachers. Responses were then anonymised at source before being shared with the evaluation team, with no identifiable school or teacher information included in the dataset.

A total of 70 teachers completed the end-point survey: 19 taught control classes only, 29 taught intervention classes only, 22 taught across both groups.

Teacher Interviews

A small qualitative component was embedded within the trial using an explanatory sequential mixed-methods design to help interpret and contextualise quantitative findings (Ivankova, Creswell, & Stick, 2006). Following analysis of pupil outcomes and a post-trial teacher survey, three semi-structured interviews were conducted with participating teachers via video conference (~30-40 minutes each). Interviews were audio-recorded with consent, professionally transcribed, and anonymised prior to analysis. Videoconference interviewing is an established and viable mode for qualitative data collection, offering pragmatic benefits with minimal loss of depth (Archibald et al., 2019). Given the evaluative purpose and narrow focus on implementation mechanisms, a very small sample was judged appropriate; in such contexts, depth and specificity can yield sufficient “information power” for analytic aims (Malterud, Siersma, & Guassora, 2016).

Analysis

Analysis combined reflexive thematic analysis to generate inductive insights with the Framework Method to structure interpretation against pre-specified constructs. Reflexive thematic analysis was first used to familiarise with the data and develop initial inductive codes and candidate themes grounded in teachers’ accounts (Braun & Clarke, 2006). The Framework Method was then applied to organise data in a case-by-theme matrix, supporting transparent comparison across interviews and alignment with the survey domains and the programme logic (Gale, Heath, Cameron, Rashid, & Redwood, 2013).

A deductive-inductive codebook underpinned integration. Deductive seeding drew on survey/logic-model domains (e.g., usability, fidelity, access, engagement), after which line-by-line coding extended the codebook inductively to capture emergent mechanisms or constraints articulated by teachers. Code definitions and indicative excerpts were iteratively refined to address overlaps between deductive and inductive categories and to structure major/minor codes.

Mixed-methods integration proceeded through two, complementary strategies. First, a joint display was developed to align quantitative results (RCT and survey) with qualitative themes, enabling “side-by-side” interpretation of patterns and mechanisms (Guetterman, Fetters, & Creswell, 2015). Second, a convergence coding matrix recorded the relationship between strands at the construct level, categorising evidence

as convergent, complementary, dissonant, or silent, to enhance analytic transparency (Farmer, Robinson, Elliott, & Eyles, 2006). The combined approach is well suited to small, practice-oriented datasets, producing auditable matrices that support cross-case comparison and theory-linked interpretation (Gale et al., 2013; Guetterman et al., 2015).

1. Deductive seeding: creating an initial code list from the survey domains and logic model.
2. Familiarisation with transcripts;
3. Inductive extension: line-by-line coding to identify inductive codes present in transcripts, adding codes where teacher language introduced new mechanisms or constraints
4. Codebook consolidation: initial deductive codes from survey domains (e.g., usability, fidelity, access, engagement) augmented inductively to capture novel issues raised by teachers; defining each code and adding indicative excerpts
5. Indexing/coding: one analyst independently coding transcripts and reviewing the codebook, updating code definitions when identifying overlap between inductive and deductive codes and grouping major/minor codes;
6. Identifying key themes: encompassing multiple major/minor codes;
7. Triangulating findings: using a convergence coding matrix and a joint display to record whether evidence converged, complemented, diverged, or was silent.

This approach is appropriate for very small samples and produces transparent matrices that facilitate comparison across distinct, diverse datasets.

Ethics and Data Protection

The trial involved only activities consistent with routine teaching. No sensitive pupil data were collected beyond attainment scores and Pupil Premium status. All data shared with evaluators were anonymised at source. Processing complied with GDPR on the basis of public interest, as the evaluation supports school and sector decision-making.

Impact Results: Summary

The final analytic sample comprised 2,296 pupils across 16 schools, randomised at the class level (74 treatment classes; 76 control classes). There were 1,204 pupils in the treatment group and 1,092 in the control group. Pupil Premium (PP) representation was comparable across groups (22.18% treatment; 21.89% control). Baseline attainment was well balanced. Mean pre-test scores were 1039.55 (SD = 73.57) in the treatment group and 1036.13 (SD = 73.03) in the control group. Post-test raw means were slightly higher in the treatment group (1060.51 vs 1053.34), though adjusted analyses account for baseline differences and clustering.

The intention-to-treat (ITT) analysis showed that pupils in the Eedi group scored 4.99 points higher on the post-test than control pupils after adjusting for baseline attainment and clustering. While the direction of effect favoured the intervention, the confidence interval included zero and the result was not statistically significant.

The interaction analysis found no statistically significant differential effect by Pupil Premium status. Effects were positive for both groups and somewhat larger for PP pupils, but estimates were not statistically significant. There was considerable heterogeneity across schools. Effect sizes were generally small and ranged in both positive and negative directions. Among pupils scoring below the median at pre-test, the adjusted mean difference was 3.90 points in favour of the intervention. This subgroup analysis did not demonstrate statistically significant impact.

Complier Average Causal Effect (CACE) analyses examined outcomes among pupils who actively engaged with Eedi. Effects were consistently larger than the ITT estimate and increased with higher engagement thresholds. Among PP pupils, effects were notably stronger and increased substantially with engagement. These findings suggest that higher engagement with the intervention is associated with greater attainment gains, particularly for Pupil Premium pupils.

Impact Results: Technical Analysis

Participant Characteristics

Not all students with pre-test scores had corresponding post-test data and vice versa. We obtained 2,515 pre-test results (1,225 from control and 1,290 from intervention students) and 2,493 post-test results (1,183 from control and 1,310 from intervention students). Including only students with matched pre- and post-test scores, the final analytic sample consisted of 2,296 students across 16 schools. Randomisation was done at the class level, with 76 and 74 classes in the control and treatment groups respectively. The number of students in the intervention group ($n = 1,204$) was higher than in the control group ($n = 1,092$). Additional baseline characteristics, including summaries stratified by Pupil Premium status, are presented in Table 1.

Characteristic	Control	Treatment
<i>Class (N)</i>	76	74
<i>All pupils (N)</i>	1092	1204
<u>Gender</u>		
<i>Male</i>	418	461
<i>Female</i>	498	545
<i>Unspecified</i>	176	198
<u>Test Scores</u>		
<i>Pretest mean (SD)</i>	1036.13 (73.03)	1039.55 (73.57)
<i>Post test mean (SD)</i>	1053.34 (78.78)	1060.51 (76.95)
<u>Pupil Premium Status</u>		
<i>PP (N)</i>	239	267
<i>non-PP(N)</i>	797	934
<i>Pupil-level (%)</i>	21.89%	22.18%

Table 1. Participant Characteristics

ITT Analysis

Students who used Eedi scored on average 4.99 (-1.65, 11.63) points higher on the post-test compared to the control group, after adjusting for baseline performance (Table 2). The corresponding hedges g effect size was 0.101 [-0.078, 0.28] (Table 3).

Parameters	Estimate	Lower	Upper
<i>Intercept</i>	113.33	-50.71	277.38
<i>Pre Score between</i>	0.90	0.75	1.06
<i>Pre-score within</i>	0.68	0.64	0.71
<i>Treatment</i>	4.99	-1.65	11.63

Table 2: Primary Analysis

Groups	Sample Size	Adjusted Mean	Cohen's d	Hedge's G	P-value
<i>Treatment</i>	1204	1055.783	0.101 (-0.041 to 0.242)	0.101 (-0.078 to 0.28)	0.2703
<i>Control</i>	1092	1050.791			

Table 3: ITT Analysis

Secondary Analysis

Pupil Premium: Interaction Analysis

To test whether the intervention had differential effects based on Pupil Premium status, an interaction term was added to the model. The analysis examined whether the intervention had a differential effect on students receiving Pupil Premium. The point estimate of the interaction suggested a somewhat larger treatment effect for Pupil Premium students; however, this difference was not statistically significant (see Table 4). Therefore, there was insufficient evidence to conclude that the intervention had a different impact on the two groups (5.92, 95% CI: -3.26 to 15.11).

Parameters	Estimate	Lower	Upper
<i>Intercept</i>	103.46	-54.30	261.22
<i>Pre Score between classes</i>	0.92	0.76	1.07
<i>Pre-score within clusters</i>	0.67	0.64	0.71
<i>Treatment</i>	3.09	-3.87	10.05
<i>Pupil Premium (PP)</i>	-14.52	-21.36	-7.68
<i>Treatment × PP</i>	5.92	-3.26	15.11

Table 4: Pupil Premium interaction analysis

Pupil Premium: Subgroup Analysis

Although the interaction analysis did not demonstrate statistically significant differential effects, a subgroup analysis was conducted for exploratory purposes (Table 5). The same model used for the primary analysis was applied separately to each Pupil Premium (PP) subgroup. Subgroup analysis showed that the effect of the intervention was positive for both premium and non-premium pupils.

Category	Groups	Sample Size	Adjusted Mean	Cohen's d	Hedges g	P-value
<i>PP</i>	<i>Treatment</i>	267	1037.064	0.178 (-0.021 to 0.376)	0.177 (-0.017 to 0.372)	0.0732
	<i>Control</i>	239	1028.288			
<i>Non-PP</i>	<i>Treatment</i>	934	1060.206	0.052 (-0.107 to 0.211)	0.052 (-0.136 to 0.24)	0.5890
	<i>Control</i>	797	1057.700			

Table 5: Pupil Premium Sub-Group Analysis

School-level Analysis

Examination of results at the individual school level revealed considerable heterogeneity, with no evidence of a statistically significant treatment effect in any school. Adjusted mean differences between treatment and control groups were generally small, and standardised effect sizes (Cohen's *d* and Hedges' *g*) were close to

zero across sites. Several schools showed small positive effects favouring the treatment group (e.g., Schools 2, B, C, I, and K; $g \approx 0.17-0.25$), while others showed small negative effects favouring the control group (e.g., Schools 1, 4, H, and E; $g \approx -0.08$ to -0.24). In many schools (e.g., Schools 3, 5, J, and L), treatment and control adjusted means were nearly identical, resulting in negligible effect sizes. All confidence intervals for effect size estimates included zero, indicating substantial uncertainty. Several schools also had relatively small sample sizes, further limiting statistical power and the precision of estimates. School-level analyses do not demonstrate a consistent or systematic advantage for either the treatment or control condition, and results should be interpreted with caution.

Category	Groups	Sample Size	Adjusted Mean	Cohen's d	Hedge's G	P-value
School 1	Treatment	98	1015.467	-0.083 (-0.368 to 0.203)	-0.082 (-0.362 to 0.197)	0.5645
	Control	98	1019.043			
School 2	Treatment	95	1073.025	0.241 (-0.316 to 0.798)	0.24 (-0.076 to 0.557)	0.1369
	Control	78	1060.817			
School 3	Treatment	64	1046.921	0.085 (-0.266 to 0.437)	0.085 (-0.261 to 0.43)	0.6307
	Control	64	1043.376			
School 4	Treatment	72	1018.7	-0.19 (-0.913 to 0.532)	-0.189 (-0.698 to 0.32)	0.4659
	Control	80	1026.68			
School 5	Treatment	31	1045.185	-0.015 (-0.68 to 0.649)	-0.015 (-0.648 to 0.618)	0.9631
	Control	16	1045.954			
School 6	Treatment	56	1088.637	0.054 (-0.836 to 0.944)	0.054 (-0.334 to 0.441)	0.7862
	Control	51	1086.091			
School A	Treatment	42	1047.328	0.092 (-0.346 to 0.53)	0.091 (-0.33 to 0.513)	0.6705
	Control	44	1043.096			

School B	Treatment	108	1070.94	0.183 (-0.211 to 0.576)	0.182 (-0.119 to 0.484)	0.2364
	Control	111	1061.842			
School C	Treatment	114	1085.163	0.246 (-0.129 to 0.62)	0.245 (-0.047 to 0.538)	0.1005
	Control	114	1076.358			
School E	Treatment	55	1125.178	-0.076 (-0.492 to 0.34)	-0.075 (-0.476 to 0.325)	0.7121
	Control	42	1128.147			
School G	Treatment	68	1112.653	0.169 (-0.67 to 1.009)	0.168 (-0.5 to 0.837)	0.6214
	Control	63	1105.024			
School H	Treatment	86	1012.999	-0.236 (-1.021 to 0.549)	-0.235 (-0.739 to 0.269)	0.3610
	Control	58	1022.22			
School I	Treatment	118	1059.682	0.17 (-0.451 to 0.792)	0.17 (-0.322 to 0.661)	0.4991
	Control	88	1050.769			
School J	Treatment	70	1057.248	0.045 (-0.306 to 0.396)	0.045 (-0.294 to 0.384)	0.7951
	Control	63	1055.232			
School K	Treatment	88	1066.369	0.232 (-0.476 to 0.939)	0.23 (-0.252 to 0.712)	0.3486
	Control	58	1055.131			
School L	Treatment	39	1011.603	0.038 (-0.572 to 0.649)	0.038 (-0.657 to 0.732)	0.9152
	Control	64	1009.99			

Table 6: School-level analysis

Students with Low Pre-Test Scores

We performed sensitivity analysis by including only subjects who scored at or below the median pre-score. After adjusting for baseline performance, students in the intervention group scored, on average, 3.90 points higher on the post-test compared

to those in the control group (95% CI: -4.84 to 12.64; see Table 7). The corresponding effect size was 0.073 (95% CI: -0.113, 0.259) as shown in Table 8.

Parameters	Estimate	Lower	Upper
<i>Intercept</i>	-629.48	-1214.01	-44.94
<i>Pretest score between</i>	1.67	1.08	2.27
<i>Pretest score within</i>	0.70	0.64	0.76
<i>Treatment</i>	3.90	-4.84	12.64

Table 7: Multilevel impact estimates for pupils with low baseline attainment

Groups	Sample Size	Adjusted Mean	Cohen d	Hedges G	P-value
<i>Treatment</i>	591	1011.322	0.073 (-0.1 to 0.246)	0.073 (-0.113 to 0.259)	0.4432
<i>Control</i>	529	1007.419			

Table 8: Adjusted post-test means and effect sizes for pupils with low baseline attainment

CACE Analysis

Table 9 presents the findings of the CACE analysis using multiple cut-offs for compliance. At all cut-offs, the effect of the intervention is positive among subjects who comply with the intervention compared to the control group. The CACE estimates are consistently larger than the ITT estimates, reflecting the concentrated effect among compliant participants. As the compliance threshold increases, the effect size grows, suggesting that higher levels of engagement with the intervention yield greater benefits.

Tables 10 and 11 present the CACE analysis findings for PP and non-PP students, respectively. Across all engagement cut-off points, the effect of Eedi was positive in both groups. Moreover, the effect size increased with higher levels of engagement, indicating that greater engagement with Eedi was associated with improved student outcomes. While the ITT analysis provides a conservative estimate of the intervention's

effectiveness in a real-world implementation scenario, the intervention shows substantial efficacy among those who actively engage with it.

Cut-off	N Compliers	% Compliers	Hedges' g (95% CI)
>=60	1020	84.70%	0.132 (-0.048, 0.312)
>=90	930	77.20%	0.145 (-0.037, 0.327)
>=120	810	67.30%	0.167 (-0.02, 0.355)
>=200	585	48.60%	0.237 (0.031, 0.443)

Table 9: CACE Analysis (Overall)

Cut-off	N Compliers	% Compliers	Hedges' g (95% CI)
>=60	201	75.30%	0.26 (0.061, 0.458)
>=90	167	62.50%	0.315 (0.108, 0.522)
>=120	140	52.40%	0.378 (0.16, 0.596)
>=200	96	36%	0.568 (0.318, 0.819)

Table 10: CACE Analysis (PP)

Cut-off	N Compliers	% Compliers	Hedges' g (95% CI)
>=60	818	87.60%	0.066 (-0.122, 0.254)
>=90	762	81.60%	0.071 (-0.118, 0.26)
>=120	670	71.70%	0.081 (-0.112, 0.273)
>=200	489	52.40%	0.113 (-0.096, 0.321)

Table 11: CACE Analysis (Non-PP)

IPE Survey Results: Summary

Seventy teachers took part in the IPE survey, providing a broad cross-section of maths specialists with substantial classroom experience, over half had taught for more than ten years. Most respondents were class teachers or middle leaders, with only a small number of senior leaders represented.

Across all conditions, teachers reported high confidence in their mathematics teaching, particularly around fostering positive attitudes, identifying learning needs, and differentiating for diverse pupils. Confidence was slightly lower, though still positive, when it came to motivating lower-attaining pupils and using technology effectively. Control-group teachers tended to report the highest confidence overall, but differences between groups were small.

Homework-setting practices were similar across groups, with most teachers assigning around four maths homework tasks per month. Tasks typically took between 25 and 30 minutes to complete, and completion rates were high, particularly in the intervention-only classes, where over half of teachers said that three-quarters or more of homework was completed. Teachers using Eedi generally found monitoring homework straightforward, with around two-thirds describing the process as “slightly” or “very straightforward.” Around a third of teachers also reported using Eedi in lessons, most commonly to track progress or use diagnostic questions, while a smaller proportion explored features such as the live tutor or Present function.

The majority of teachers (82%) attended Eedi training and found the platform easy to set up and navigate. Implementation barriers were primarily external rather than platform-specific. Student motivation (72%) and limited device or internet access at home (71%) were the most common challenges, followed by curriculum misalignment (61%). In contrast, few teachers reported that Eedi increased workload, and most found lesson planning and homework management no more time-consuming than usual practice. Teachers delivering the intervention typically spent more time preparing lessons than those teaching control classes, though many described workload as stable or even reduced.

Teachers’ views on engagement were mixed but generally positive. Around a third of intervention-only teachers felt that pupil engagement had increased, and a similar proportion saw no change. Dual-group teachers were more uncertain, perhaps reflecting the complexity of running both conditions simultaneously. Most felt that disadvantaged pupils (PP/FSM) required no additional support to use Eedi, though

around one in three thought these pupils needed slightly more help with access or navigation.

Feedback on Eedi's quiz design highlighted both enthusiasm and constructive suggestions. Teachers praised the ease of setting weekly homework and the potential for retrieval practice but felt that some quizzes were too difficult for weaker students, too short to reflect normal homework expectations, or not always well aligned with their scheme of work. Requests for improvement included more consistent question difficulty, longer or customisable quizzes, the ability to assign multiple quizzes at once, and better topic coverage.

Overall, teachers viewed Eedi as a useful, time-saving tool that supported homework setting and progress monitoring without adding substantially to workload. The platform was considered easy to use and positively received, particularly where it aligned well with the curriculum and pupils could access it reliably. Where challenges remained, they related less to the platform itself and more to motivation, access, and curriculum fit.

IPE Survey Results: Technical Analysis

Teacher Characteristics

70 IPE responses were obtained from teachers. 19 teachers taught control group students only, 29 teachers taught intervention group students only and 21 teachers taught across both intervention and control pupil groups (dual-group teachers). Across groups, the sample was predominantly mathematics specialists with a broad range of teaching experience. Of the 19 control-group-only teachers: 1 was a senior leader, 5 head of department, 3 assistant head of department, 10 class teachers. All but one identified as specialist maths teachers, with the remaining teacher stating their specialism was Physical Education. Of the 29 intervention-group-only teachers: 1 was a senior leader, 4 head of department, 3 assistant head of department, 21 class teachers. All identified as specialist maths teachers. Of the dual-group teachers, there were no senior leaders, 1 head of department, 3 assistant head of department, 16 class teachers and 1 supply teacher. All identified as specialist maths teachers, apart from one whose specialism was Physical Education. Teaching experience was extensive across all groups: 57% of teachers had 10 or more years' experience.

Teaching Condition	Years Teaching Experience	N	%
<i>Control-Group-Only</i>	1-2	1	5%
	3-5	4	21%
	6-10	5	26%
	10+	9	47%
<i>Intervention-Group-Only</i>	1-2	3	10%
	3-5	3	10%
	6-10	6	21%
	10+	17	59%
<i>Dual-Group</i>	1-2	4	18%
	3-5	0	0%
	6-10	4	18%
	10+	14	64%

Table 12. Teacher characteristics

Teacher Confidence

A scale adapted from Hettinger et al., (2022) was used to explore teacher confidence in teaching mathematics. Overall, confidence levels were high across all domains and groups, with most teachers reporting that they felt slightly or very certain/confident about their practice. Control-only teachers tended to show the highest confidence overall, particularly in helping pupils value learning in maths and identifying students needing support. Confidence was strongest for relational and diagnostic aspects of teaching such as fostering positive attitudes and identifying learning needs. It was slightly lower, though still positive, for motivating lower-attaining pupils and using technology effectively.

Helping pupils value learning in maths

Across all groups, teachers expressed strong confidence in their ability to help pupils value learning in mathematics. The majority reported feeling 'very certain' (ranging from 59% to 79% across groups), with an additional 16-36% feeling 'slightly certain'. Only a small minority expressed uncertainty or neutrality (7% or fewer in each group). Control-only teachers showed the highest overall confidence, with 79% 'very certain' compared to 72% in the intervention-only group and 59% in the dual-group cohort.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control-Only (N)</i>	0	0	1	3	15
<i>Control-Only (%)</i>	0%	0%	5%	16%	79%
<i>Intervention-Only (N)</i>	2	0	0	6	21
<i>Intervention-Only (%)</i>	7%	0%	0%	21%	72%
<i>Dual-Group (N)</i>	1	0	0	8	13
<i>Dual-Group (%)</i>	5%	0%	0%	36%	59%

Table 13. Responses to How certain are you that you can help pupils value learning in maths?

Motivating students with low interest in maths

Teachers generally expressed moderate to high confidence in their ability to motivate students with low interest in mathematics. Across all groups, 5% or fewer reported feeling uncertain, with a further 3-11% neutral.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control-Only (N)</i>	0	1	2	8	8
<i>Control-Only (%)</i>	0%	5%	11%	42%	42%
<i>Intervention-Only (N)</i>	0	1	1	16	11
<i>Intervention-Only (%)</i>	0%	3%	3%	55%	38%
<i>Dual-Group (N)</i>	1	0	2	13	6
<i>Dual-Group (%)</i>	5%	0%	9%	59%	27%

Table 14. How certain are you that you can motivate students who show low interest in maths?

Motivating students who often fail school requirements

The majority of teachers felt '*slightly certain*' about their ability to motivate students who often fail school requirements, but confidence levels varied across groups.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control-Only (N)</i>	0	2	5	9	3
<i>Control-Only (%)</i>	0%	11%	26%	47%	16%
<i>Intervention- Only (N)</i>	0	0	7	15	7
<i>Intervention-Only (%)</i>	0%	0%	24%	52%	24%
<i>Dual-Group (N)</i>	1	0	8	13	0
<i>Dual-Group (%)</i>	5%	0%	36%	59%	0%

Table 15. How certain are you that you can motivate even those students who often fail school requirements?

Identifying students who need support across maths topics

Teachers across all groups expressed high confidence in identifying which students need support across different mathematical topics. The majority of control-only teachers (74%) reported feeling 'very certain', with smaller proportions 'slightly certain' (16%) or 'neutral' (5%). Confidence was similarly strong among intervention-only and dual-group teachers, with around half in each group 'slightly certain' (52% and 50%, respectively) and a further 45% 'very certain'. Only one teacher, in the control-only group (5%), reported feeling 'slightly uncertain'.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control-Only (N)</i>	0	1	1	3	14
<i>Control-Only (%)</i>	0%	5%	5%	16%	74%
<i>Intervention-Only (N)</i>	0	0	1	15	13
<i>Intervention-Only (%)</i>	0%	0%	3%	52%	45%
<i>Dual-Group (N)</i>	0	0	1	11	10
<i>Dual-Group (%)</i>	0%	0%	5%	50%	45%

Table 16. How certain are you that you can identify which students need support across different topics in maths?

Differentiating to meet diverse needs in maths

Teachers reported strong confidence in their ability to differentiate for diverse learning needs in maths. Over four in ten respondents across all groups indicated they were 'very certain' (ranging from 35% to 45%), while around half reported being 'slightly certain' (41-55%). Only a small minority were 'neutral' or 'slightly uncertain' (5-14%), and no teachers selected very uncertain.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control-Only (N)</i>	0	0	1	10	8
<i>Control-Only (%)</i>	0%	0%	5%	53%	42%
<i>Intervention-Only (N)</i>	0	0	4	12	13
<i>Intervention-Only (%)</i>	0%	0%	14%	41%	45%
<i>Dual-Group (N)</i>	0	1	1	12	8
<i>Dual-Group (%)</i>	0%	5%	5%	55%	35%

Table 17. How certain are you that you can differentiate to meet the diverse needs of your students in maths?

Using technology to enhance maths learning

Across all groups, teachers expressed strong confidence that using technology can enhance student learning in mathematics. Nearly half of control-only (47%) and intervention-only (48%) teachers described themselves as 'very confident', with a further 31-45% 'slightly confident'. Confidence was slightly lower among dual-group teachers.

Condition	Very Unconfident	Slightly Unconfident	Neutral	Slightly Confident	Very Confident
<i>Control-Only (N)</i>	0	1	1	8	9
<i>Control-Only (%)</i>	0%	5%	5%	42%	47%
<i>Intervention-Only (N)</i>	0	2	4	9	14
<i>Intervention-Only (%)</i>	0%	7%	14%	31%	48%
<i>Dual-Group (N)</i>	0	2	2	10	8
<i>Dual-Group (%)</i>	0%	9%	9%	45%	36%

Table 18. How confident are you that using technology can enhance student learning in maths?

Maths Homework

Time Expectations and Completion Rates

Of the 19 control-only teachers, 18 reported setting homework for their students. On average, pupils were assigned around four pieces of maths homework per month, ranging from 1-12 pieces, with pupils typically expected to spend 32 minutes per task (range 15-60 minutes).

Of the 29 intervention-only teachers, 27 reported setting maths homework. On average, students received four pieces of homework per month, ranging from 1-12 pieces, with an average expected completion time of 26 minutes per task (range 3-60 minutes).

Among teachers who taught both Intervention and Control groups (n=22), 20 reported setting homework with their intervention classes. On average, these teachers assigned four pieces of homework per month, ranging from 2-16 pieces, with pupils typically expected to spend 24 minutes per task (range 8-40 minutes). 15 reported setting homework with their control classes. On average, control classes were assigned 4 pieces of homework per month, with pupils typically expected to spend 28 minutes per task (range 5-60 minutes).

Homework completion was generally high, with most teachers reporting that pupils completed at least half of assigned tasks. The highest completion rates were reported by intervention-only teachers, with 56% indicating that over three-quarters of homework was completed. Control-only teachers showed a similar trend (44%), while dual-group classes, particularly control classes, reported more variable completion, with a greater proportion of teachers (up to 37%) indicating lower engagement levels.

Condition	Completion Rate	N	%
<i>Intervention-Only</i>	< 25%	3	10%
	25-50%	2	7%
	50-75%	7	26%
	75%+	15	56%
<i>Control-Only</i>	< 25%	2	11%
	25-50%	1	6%
	50-75%	7	39%
	75%+	8	44%

<i>Dual-Group (Intervention)</i>	< 25%	5	23%
	25-50%	4	18%
	50-75%	4	18%
	75%+	7	32%
<i>Dual-Group (Control)</i>	< 25%	3	14%
	25-50%	0	0%
	50-75%	7	32%
	75%+	5	23%

Table 19. Homework completion rate

Platform Use

12 control-group-only teachers reported regularly using an online platform to assign maths homework (Dr Frost n=2, Mathswatch n=1, Sparx n=8, MyMaths, n=1, Mathspad n=1). 22 intervention-group-only teachers reported using an online platform to assign maths homework; 14 were using the intervention platform, Eedi. Of those using Eedi, 10 also used additional platforms (Dr Frost n=5, Mathswatch n=5, Sparx n=2, Mathspad n=2). 7 of the 12 control-group-only teachers indicated how frequently these platforms were used: 1 set online platform homework daily, 1 more than twice a week, 1 twice weekly, 2 weekly, 1 fortnightly, 1 less than monthly.

Eight of the twenty-nine intervention-group-only teachers were not using Eedi to assign homework and instead used alternative platforms (Dr Frost n=1, Mathswatch n=1, Sparx n=7).

18 dual-group teachers reported using online systems to set maths homework for intervention classes, 11 of which were using Eedi. Of those using Eedi, 8 also used additional platforms (Dr Frost n=5, Mathswatch n=1, Sparx n=3, MyMaths n=1, HegartyMatches n=1, Corbett n=1, MathsPad n=1). Seven dual-group teachers were not using Eedi to assign homework with intervention classes, but were all using alternative platforms (Dr Frost n=2, Mathswatch n=1, Sparx n=6, MathsPad n=1, LBQ n=1).

11 dual-group teachers reported using online systems to set maths homework for control classes (Dr Frost n=3, Mathswatch n=2, Sparx n=6, MathsPad n=2). Seven dual-group teachers were not using Eedi to assign homework with intervention classes but were all using alternative platforms (Dr Frost n=2, Mathswatch n=1, Sparx n=6,

MathsPad n=1, LBQ n=1). 8 of the 11 control-group-only indicated how frequently these platforms were used: 1 more than twice a week, 1 weekly, 3 fortnightly, 1 monthly, 2 less than monthly.

Monitoring homework via Eedi - Teacher experience

Teachers generally found monitoring homework through *Eedi* to be a clear and manageable process. The majority of respondents across both groups rated the experience as slightly or very straightforward (58% of intervention-only and 61% of dual-group teachers).

Condition	Rating	N	%
<i>Intervention-Only</i>	Very Challenging	1	5%
	Slightly Challenging	6	32%
	Neutral	1	5%
	Slightly Straightforward	7	37%
	Very Straightforward	4	21%
<i>Dual-Group (Intervention)</i>	Very Challenging	1	8%
	Slightly Challenging	3	23%
	Neutral	1	8%
	Slightly Straightforward	5	38%
	Very Straightforward	3	23%

Table 20. Teachers' experiences of monitoring homework via Eedi

Monitoring homework - control group teacher experience

Across both control and dual-group control conditions, teachers generally found monitoring homework to be straightforward. Over half of control-only teachers (56%) and a third of dual-group control teachers (33%) described the process as 'very straightforward', while an additional 11% and 47%, respectively, found it 'slightly straightforward'.

Condition	Rating	N	%
<i>Control Only</i>	Very Challenging	0	0%
	Slightly Challenging	4	22%
	Neutral	2	11%
	Slightly Straightforward	2	11%
	Very Straightforward	10	56%
<i>Dual-Group (Control)</i>	Very Challenging	0	0%
	Slightly Challenging	1	7%
	Neutral	2	13%
	Slightly Straightforward	7	47%
	Very Straightforward	5	33%

Table 21. Control group teachers' experiences of monitoring homework

Eedi Classwork

Ten of the twenty-nine Intervention-Only teachers (34%) reported using Eedi during lessons, while the remaining nineteen (66%) did not. Among teachers who taught both Intervention and Control groups, nine out of twenty-two (41%) reported using Eedi in lessons, with the remaining thirteen (59%) not using it during class time.

Monitoring classwork completion and performance

Among Intervention-Only teachers using Eedi in class, perceptions of monitoring were generally positive, with the majority (56%) describing the process of tracking classwork completion and performance as 'Slightly Straightforward' and a further 22% as 'Very Straightforward'. Among Dual-Group (Intervention) teachers, views were more mixed but still leaned positive. Nearly half (44%) rated the process as 'Slightly Straightforward', with 22% describing it as 'Slightly Challenging'.

Condition	Rating	N	%
<i>Intervention Only</i>	Very Challenging	0	0%
	Slightly Challenging	1	11%
	Neutral	2	22%

	Slightly Straightforward	5	56%
	Very Straightforward	2	22%
<i>Dual-Group (Intervention)</i>	Very Challenging	0	0%
	Slightly Challenging	2	22%
	Neutral	3	33%
	Slightly Straightforward	4	44%
	Very Straightforward	0	0%

Table 22. Teacher experiences of using Eedi to monitor classwork

Eedi Functions

To understand how the platform was integrated in lessons, teachers using Eedi in class time with intervention classes (n=19, including intervention-only and dual-group teachers) were asked if they had used features including Present, diagnostic questions, Eedi resources, and the live human tutor.

Present function

Use of the Present function was mixed. Three teachers reported using the feature in class to discuss misconceptions with students, while sixteen (84%) did not. Among the sixteen non-users, eleven (69%) were unaware of the feature, and five (31%) reported knowing about it but choosing not to use it.

Diagnostic Questions

Of the 19 teachers who used Eedi in class, thirteen (68%) reported using Eedi's diagnostic questions during lessons, while six (32%) did not.

Eedi Resources

Three teachers (16%) reported using supplementary materials such as videos or worksheets in their classroom teaching, while sixteen (84%) did not.

Live Human Tutor

Use of the live human tutor feature on Eedi was moderate. Four teachers (25%) reported not using the feature enough to detect an impact on practice, while twelve (75%) reported using it and provided feedback on their experience. Among the twelve who used the live tutor: one teacher (8%) felt it enhanced differentiated instruction capabilities, five (42%) reported that it improved student independence in learning,

two (17%) said it reduced time spent on individual student questions, and four (33%) reported no noticeable impact on their teaching. In terms of overall perceptions of usefulness, one teacher (6%) found the live tutor very unhelpful, one (6%) somewhat unhelpful, three (19%) neither helpful nor unhelpful, seven (44%) somewhat helpful, and three (19%) very helpful. Teachers were also asked to provide feedback on their experiences using the live tutor. Three free-text responses were submitted, highlighting issues with responsiveness and workflow disruption when students engaged with the live tutor.

1	Children were waiting while one person helped multiple people. They were unable to continue with the questions while in a conversation with the human
2	It stopped my students from answering the question once they find their answer.
3	Some tutors did not get back to the students quickly enough

Table 23. Free-text responses on the Live Tutor

Lesson Use

To understand how Eedi was used in lessons, intervention teachers reported whether they had: reviewed common wrong answers with the whole class; used insights to inform planning for the next lesson; created small groups for targeted support; identified individual student misconceptions; shared student responses to facilitate class discussion; and set quizzes.

Reviewing common wrong answers with the whole class

Two teachers (4%) reported using *Eedi* to review common wrong answers with the whole class almost every lesson, eight (16%) did so weekly, another eight (16%) monthly, twenty (39%) rarely, and thirteen (25%) never used it for this purpose.

Informing planning for the next lesson

In terms of using *Eedi* to inform planning for the next lesson, one teacher (2%) reported doing so almost every lesson, another one (2%) every lesson, ten (20%) weekly, seven (14%) monthly, twenty (39%) rarely, and twelve (24%) never.

Creating small groups for targeted support

In relation to using *Eedi* to create small groups for targeted support, one teacher (2%) reported doing so almost every lesson, two (4%) weekly, two (4%) monthly, nineteen (37%) rarely, and twenty-seven (53%) never.

Identifying individual student misconceptions

When asked about using Eedi to identify individual student misconceptions, one teacher (2%) reported doing so almost every lesson, ten (20%) weekly, thirteen (25%) monthly, thirteen (25%) rarely, and fourteen (28%) never.

Sharing student responses to facilitate class discussion

Regarding the use of Eedi to share student responses and facilitate class discussion, six teachers (12%) reported doing so weekly, five (10%) monthly, fifteen (29%) rarely, and twenty-five (49%) never.

Quizzes

Eleven teachers did not use Eedi quizzes. Among the forty teachers who provided feedback on curriculum alignment, three (8%) reported no alignment between quizzes and the curriculum, ten (25%) indicated little alignment, twenty-three (58%) noted moderate alignment, and four (10%) described complete alignment.

Twenty teachers (39%) found retrieval quizzes most beneficial, twenty-one (41%) preferred topic quizzes, and ten (20%) were unsure which quiz type was most useful.

The Eedi programme aimed for teachers to set approximately one quiz per week for their students. Against this target, twenty-four teachers (47%) reported consistently setting one or more quizzes weekly, while twenty-one (41%) said they aimed to do so but often set fewer. A smaller group of six teachers (12%) reported intentionally setting fewer quizzes.

Teachers identified several factors that limited their ability to set Eedi quizzes regularly. The most common reasons were a preference for other tools or approaches, with eighteen teachers (35%) reporting they preferred to use other digital platforms or paper-based homework. Thirteen (25%) said they did not have enough time for the planning and administration involved, and twelve (24%) felt the quizzes did not align well enough with their curriculum pacing. Eleven teachers (22%) noted low student engagement or completion rates, and ten (20%) cited concerns about students' access to devices or the internet at home. Smaller proportions reported encountering technical issues with the Eedi platform (8%, n=4), not finding the data and insights useful enough to justify the time (10%, n=5), or that school or departmental priorities did not support weekly quiz-setting (16%, n=8).

Teachers were asked to share their thoughts on Eedi's quiz types. Seventeen open-ended responses were submitted.

1	The quizzes were very difficult for the bottom set. It would be nice to have quizzes where all 5 questions are very similar. I have found that usually all 5 questions are very different, so when students need specific practice, they can't do that on Eedi.
2	Sometimes there were technical difficulties with the system
3	I could not find quizzes on every topic on our SOL.
4	I found that the work was too "bitty" and we already use too many digital platforms. For a low set this was another log in for them to learn and was too complicated.
5	Quite a high level for very weak students
6	Easy to set weekly homework
7	Sorry, I was the second teacher on this class (minority teacher) so I was not the teacher who set the home learning etc.
8	A way to add more than one quiz at a time, and for the search engine to retain/remember previously searched topics
9	The front end is awful
10	It is really positive effect on the students willingness to do homework
11	Allowing teacher to use questions difficulty
12	They need to be longer to align with expected homework completion times
13	I found that sometimes when a student puts their answer in and it is correct, the question has either marked it as wrong or wrote in a different format that the student isn't used to.
14	Retention quizzes were good, but generally my class being on the weaker side struggled with these.
15	Not too popular with students
16	Quizzes sometimes varied in difficulty level. Occasionally much too difficult for the assigned age group.
17	I think 10 questions would be much better than 5 as my class complete the tasks very quickly

Table 24. Free text responses on Eedi quizzes

Responses offer a mix of positive feedback and constructive suggestions for improvement.

Appropriateness of Difficulty and Differentiation

A central theme across responses was concern about quiz difficulty, especially for lower-attaining or "bottom set" pupils. Several teachers felt the questions were pitched too high, too varied, or insufficiently scaffolded.

"The quizzes were very difficult for the bottom set. It would be nice to have quizzes where all 5 questions are very similar."

"Quite a high level for very weak students"

"Retention quizzes were good, but generally my class being on the weaker side struggled with these."

Teachers also observed inconsistency in difficulty within and across quizzes: *"Quizzes sometimes varied in difficulty level. Occasionally much too difficult for the assigned age group."* This indicates that while Eedi's adaptive and diagnostic structure may have worked well for some, others wanted greater control over difficulty and progression, including an option for teachers to set or filter by level: *"Allowing teacher to use questions difficulty."*

Alignment with Curriculum and Content Coverage

A number of teachers highlighted issues of curriculum alignment and content availability.

"I could not find quizzes on every topic on our scheme of learning,"

"They need to be longer to align with expected homework completion times."

This reflects a broader theme seen elsewhere in the data: teachers found Eedi useful for homework and retrieval but sometimes struggled to fit its content precisely within their own curriculum structures.

Practical Usability and Workflow

Several comments related to ease of use, navigation, and functionality. Teachers requested interface and workflow improvements, such as *"a way to add more than one quiz at a time, and for the search engine to retain/remember previously searched topics."* Technical reliability also featured:

"Sometimes there were technical difficulties with the system,"

"I found that sometimes when a student puts their answer in and it is correct, the question has either marked it as wrong or wrote in a different format that the student isn't used to."

Student Experience and Motivation

Teacher comments reflected mixed pupil responses to Eedi quizzes. One teacher noted a *"really positive effect on the students' willingness to do homework,"* while another described Eedi as *"easy to set weekly homework."* However, other teachers found engagement uneven, particularly among lower sets: *"Not too popular with students,"* and *"For a low set this was another log in for them to learn and was too complicated."*

This suggests that while Eedi was accessible and motivating for some, others found it cognitively or practically demanding.

Suggestions for Improvement and Development Ideas

Several teachers offered practical suggestions to enhance the platform's functionality and alignment with classroom practice. These included: expanding quiz length:

"I think 10 questions would be much better than 5 as my class complete the tasks very quickly."

Workload

Time spent on lesson planning/preparation and homework

Teachers teaching intervention-only groups were asked to estimate on average, how many minutes they spend on lesson planning/preparation per week. Teachers reported highly variable weekly planning time, ranging from 3-600 minutes, with an average of 152 minutes per week. Dual-group teachers were also asked to estimate on average, how many minutes they spend on lesson planning/preparation per week for their intervention classes. Teachers reported highly variable weekly planning time, ranging from 2-960 minutes, with an average of 249 minutes per week.

Control-only teachers also reported wide variability, from 5-1200 minutes, averaging 58 minutes per week. Dual-group teachers reported spending 6-300 minutes on lesson planning/preparation per week for control classes, averaging 133 minutes.

This suggests that teachers delivering the intervention spent considerably more time on lesson planning and preparation than those teaching control classes; both among teachers teaching only intervention groups and those teaching both intervention and control groups ("dual-group" teachers).

Teachers were also asked to estimate on average how many minutes they spend on homework setting/marking per week. Intervention-group-only teachers estimated 1-180 minutes per week on homework tasks, with an average of 56 minutes. Dual-group teachers were also asked to estimate on average, how many minutes they spend on homework setting/marking per week for their intervention classes. Dual-group teachers estimated 1-360 minutes per week on homework tasks for their intervention classes, with an average of 49 minutes.

Control-group-only teachers reported 5-120 minutes per week, averaging 34 minutes. Dual-group teachers reported 5-60 minutes spent on setting/marking homework for control classes, averaging 22 minutes.

This indicates that intervention teachers generally spent more time on homework-related activities than control teachers, though this difference was smaller than that reported for lesson planning. This reinforces the pattern observed for lesson planning and could suggest that intervention delivery introduced modest additional workload, particularly for teachers running both versions of the programme. It could also be the case however that these differences between groups existed prior to the trial.

Perceived impact of Eedi on workload

Most teachers reported that using Eedi had little or no effect on their overall workload. Almost half of intervention-only teachers (45%) and half of dual-group intervention teachers (50%) reported no change. Among those who did perceive a change, a larger proportion reported decreases (44% of intervention-only; 27% of dual-group) than increases, suggesting that for many, the platform may have streamlined elements of planning or marking. One in ten intervention-only teachers (10%) and around a quarter of dual-group intervention teachers (23%) reported increased workload; Eedi was generally seen as manageable or workload-neutral by most teachers.

Condition	Reported Impact	N	%
<i>Intervention Only</i>	Significantly decreased	3	10%
	Somewhat decreased	10	34%
	Somewhat increased	2	7%
	Significantly increased	1	3%
	Neither increased or decreased	13	45%
<i>Dual-group (Intervention)</i>	Significantly decreased	0	0%
	Somewhat decreased	6	27%
	Somewhat increased	5	23%
	Significantly increased	0	0%
	Neither increased or decreased	11	50%

Table 25. Perceived impact of Eedi on teacher workload

Most teachers perceived time savings rather than added workload when using Eedi across several key areas, particularly marking and lesson preparation, though patterns varied slightly between intervention-only and dual-group teachers.

For intervention-only teachers, reductions were most commonly reported for marking or correcting pupils' work (12 teachers), followed by lesson preparation (6 teachers) and organising resources (6 teachers). Very few reported increases in these areas suggests that the intervention may have streamlined feedback and resource management processes. Time spent on administration and assessment also tended to decrease slightly overall.

For dual-group teachers, the pattern was more mixed. While some still reported decreased time on marking (6 teachers) and lesson preparation (4 teachers), others experienced increases in administrative work (5 teachers) and assessment tasks (3 teachers). This may reflect the additional coordination required to manage both intervention and control classes simultaneously.

Condition	Task	Decreased time (N)	Increased time (N)
<i>Intervention Only</i>	Individual planning or preparation of lessons	6	1
	Marking/correcting pupils work	12	0
	Planning, administering and reporting on pupil assessments	4	1
	Pupil supervision and tuition outside of timetabled lessons	3	2
	General administrative work	5	1
	Organising resources	6	1
	Communication and cooperation with parents or guardians	2	1
	Teamwork and dialogue with colleagues within this school	1	2
	Staff meetings	0	0
<i>Dual-group (Intervention)</i>	Individual planning or preparation of lessons	4	1
	Marking/correcting pupils work	6	0

	Planning, administering and reporting on pupil assessments	1	3
	Pupil supervision and tuition outside of timetabled lessons	2	1
	General administrative work	4	5
	Organising resources	1	1
	Communication and cooperation with parents or guardians	0	2
	Teamwork and dialogue with colleagues within this school	0	1
	Staff meetings	0	1

Table 26. Workload increase/decrease associated with Eedi use

Pupil Engagement

Among intervention-only teachers, just over one-third (31%) reported increases in pupil engagement, while a similar proportion (34%) saw no change. Only around one in five (20%) felt engagement had decreased, suggesting that the majority viewed Eedi as having a neutral or positive influence on pupil engagement.

For dual-group teachers, perceptions were more divided: about 18% reported decreases, 18% reported increases, and nearly a quarter (23%) were unsure. Half of this group either saw no change or didn't know, implying a more cautious or uncertain view when balancing intervention and control teaching.

Condition	Pupil Engagement	N	%
<i>Intervention Only</i>	Significantly decreased	3	10%
	Decreased	3	10%
	Increased	6	21%
	Significantly increased	3	10%
	Neither increased or decreased	10	34%
	Don't know	4	14%
<i>Dual-Group (Intervention)</i>	Significantly decreased	2	9%

	Decreased	4	18%
	Increased	2	9%
	Significantly increased	2	9%
	Neither increased or decreased	7	32%
	Don't know	5	23%

Table 27. Pupil engagement increase/decrease associated with Eedi use

When asked whether Pupil Premium (PP) and/or Free School Meal (FSM) pupils required more support than their peers to use Eedi, teachers generally reported little difference in support needs. While around one-third of teachers observed PP/FSM pupils needing slightly more help to use Eedi, the majority saw little or no difference or were unsure. Across all teachers delivering intervention classes, 18 teachers (around 38%) felt that PP/FSM pupils required about the same amount of support as their peers, while 13 teachers (27%) believed these pupils needed slightly more support. Only 2 teachers (4%) said PP/FSM pupils required significantly more support, and 1 teacher (2%) felt they needed slightly less. A further 17 teachers (36%) said they did not know, suggesting that differences may not be consistently visible in classroom practice.

Among intervention-group-only teachers, roughly equal proportions perceived no difference (11 teachers) or were unsure (11 teachers). A smaller group of 5 teachers felt that PP/FSM pupils needed slightly more support. Among dual-group teachers (those teaching both intervention and control classes), responses were similarly mixed. 7 teachers reported that PP/FSM pupils needed the same amount of support, 8 said they needed slightly more, and 1 said they needed significantly more. This group showed a slightly higher proportion perceiving additional support needs, possibly reflecting greater opportunities for direct comparison between groups.

Thirteen free-text comments were provided by teachers when asked for further comments.

1	<i>Mixed class and there is no major distinguishing difference between the students</i>
2	<i>I achieved a better completion rate with the control class I completed paper homework with.</i>
3	<i>The questions are not aligned with our sow which confused students sometimes</i>

4	I used EEDI a lot at the start of the academic year but this tail of as student engagement became low. This was due to the lack of ICT resources that students had at home, not EEDI itself. A high proportion of my learners were unable to access their home learning.
5	I have set homework on Eedi each week, but they have a poor completion rate. I have a very week nurture group who have a lot of behavioural problems. I can't compare to previous year groups as we haven't had this cohort in school prior to this year.
6	I was supportive of this being used within the department and I am very intersted to see if it has had impact.
7	I have not used Eedi this year as I have not been timetabled in a computer room, I am not the teacher responsible for setting the year 7 homeworks and my timetable has changed a few times over the year so I am teaching different students.
8	Eedi reduces my planning time because I don't have to worry about printing and marking homework for that class. But if there is a way to link the school's curriculum to eedi, and then the platform can suggest quizzes to set, that will be amazing.
9	Weaker students struggle to log in, especially at home
10	Pupils often don't read the tutor mode in detail
11	I find the students become more off topic when they have an Eedi lesson.
12	Some learner who were reluctant to write in class definite increased engagment with Eedi
13	Most of my kids forgot that they had to do eedi despite reminding. They found it annoying that they had to complete both eedi and sparx- despite the time on sparx being drastically lowered.

Table 28. Teacher free-text responses on student engagement with Eedi

Teachers' reflections on using Eedi revealed a nuanced picture of how the platform was experienced in real classrooms, shaped by access, curriculum alignment, and pupil needs.

Several teachers described how pupil engagement with Eedi was highly dependent on the resources available to them outside school. One teacher explained that they "used Eedi a lot at the start of the academic year but this tailed off as student engagement became low," adding that this was "due to the lack of ICT resources that students had at home, not Eedi itself." Similarly, another noted that "weaker students struggle to log in, especially at home," suggesting that digital access and confidence were ongoing barriers. Behavioural and contextual factors also played a part. A teacher working with a nurture group commented that they had "set homework on Eedi each week, but they have a poor completion rate," explaining that "I have a very weak nurture group who have a lot of behavioural problems." For some pupils, however, digital delivery was a strength rather than a limitation: one teacher observed that "some learners who were reluctant to write in class definitely increased engagement with Eedi." The diversity of

these responses shows that engagement was not simply a matter of liking or disliking the platform, it was closely tied to pupils' circumstances, confidence, and access to technology.

Curriculum alignment and logistical factors also strongly influenced teachers' experiences. One teacher noted that "the questions are not aligned with our scheme of work which confused students sometimes," pointing to a structural mismatch between the school's planning sequence and the content within *Eedi*. Another teacher was unable to use the platform at all because they "were not timetabled in a computer room" and had experienced timetable changes that meant they were "teaching different students" across the year. Others saw potential for improvement: "Eedi reduces my planning time because I don't have to worry about printing and marking homework for that class," one teacher explained, "but if there is a way to link the school's curriculum to Eedi, and then the platform can suggest quizzes to set, that will be amazing." These reflections suggest that while teachers valued the platform's functionality and timesaving features, they also wanted greater integration with existing schemes of work and fewer practical barriers to consistent use.

Teachers' perceptions of classroom impact varied widely. For some, there was "no major distinguishing difference between the students" when comparing groups, while others found that "I achieved a better completion rate with the control class I completed paper homework with." One teacher commented that "students become more off topic when they have an Eedi lesson," hinting at possible challenges with classroom focus during technology-based activities. Yet others described clear benefits. These contrasting accounts reflect how digital tools can both enhance and complicate classroom dynamics, boosting motivation for some learners while distracting others.

Implementation

Training Attendance

Of the 51 teachers delivering the intervention (including both intervention-only and dual-group teachers), 42 teachers (82%) reported attending an Eedi training event, while 3 teachers (6%) were unsure and 6 teachers (12%) did not attend.

Set-Up Experience

Overall, teachers found Eedi accessible and easy to use. A clear majority (n=25, 49%) described the platform as either somewhat or very straightforward, while 35% (n=18) felt it was neither challenging nor straightforward, suggesting a neutral experience for

around one-third of users. Only 16% (n=8) rated it as somewhat challenging, and none found it very challenging.

Rating	N	%
<i>Very Challenging</i>	0	0%
<i>Somewhat Challenging</i>	8	16%
<i>Neither Challenging nor Straightforward</i>	18	35%
<i>Somewhat Straightforward</i>	16	31%
<i>Very Straightforward</i>	9	18%

Table 29. Teacher responses related to experiences of Eedi set-up

Reported Barriers to Implementation

Teachers rated potential barriers to the implementation of Eedi across six areas.

Barrier	Significant Barrier N (%)	Minor Barrier (N, %)	Not a Barrier (N, %)
<i>Time required for lesson planning with Eedi</i>	4 8%	13 25%	34 67%
<i>Lack of alignment between Eedi quizzes and my curriculum sequence</i>	11 22%	20 39%	20 39%
<i>Student lack of access to devices or internet at home</i>	10 20%	26 51%	15 29%
<i>Unreliable internet or IT equipment within school</i>	8 16%	17 33%	26 51%
<i>Lack of student motivation to complete Eedi assignments</i>	15 29%	22 43%	14 27%
<i>Difficulty interpreting student results/analytics provided by Eedi</i>	8 16%	21 41%	22 43%

Table 30. Reported barriers to implementation of Eedi

Student motivation and digital access emerged as the most common barriers. Nearly three-quarters of teachers (72%) reported that a lack of student motivation was either a significant or minor barrier, making it the most frequently cited issue. Similarly, 71% of teachers indicated that limited access to devices or internet at home posed at least a minor barrier, highlighting persistent digital divide challenges. Curriculum alignment was also a notable concern: 61% of teachers viewed misalignment between Eedi quizzes and their curriculum sequence as a barrier, with 22% rating it as significant.

By contrast, fewer teachers felt that Eedi substantially increased their workload. Two-thirds (67%) said lesson planning time was not a barrier, suggesting that the platform generally did not add significant planning burden. Similarly, unreliable IT within school and difficulty interpreting analytics were viewed as minor or non-barriers by most respondents, with only a small proportion (16%) reporting these as significant obstacles.

When asked about the most common IT or technical issues faced by students when using Eedi, 20 teachers (39%) reported problems with lack of devices or internet connectivity at home, making this the most frequently cited issue. 11 teachers (22%) mentioned limited device access or unreliable connectivity at school, while 12 teachers (24%) identified platform-related issues, such as software bugs or system breakdowns. 7 teachers (14%) reported no technical issues at all.

Teachers also had the opportunity to identify further barriers through a free-text response. 1 comment was submitted: 'Lack of motivation - already using Sparx'.

IPE Interview Results

In addition to completing surveys, 3 intervention group teachers participated in semi-structured interviews (videocall; ~30-40 minutes each) to explore their experiences of and views on Eedi in greater depth. Teacher interviews were first subject to thematic analysis to identify emergent themes and then analysed using Framework Analysis with a deductive–inductive codebook as part of an explanatory sequential mixed-methods design. Insights from interviews were integrated with the RCT and survey results using a joint display and a convergence coding matrix.

Thematic Analysis

Teacher interviews were first subject to thematic analysis. We followed a classic, iterative process: (1) close reading of each transcript; (2) initial (open) coding of meaningful segments in vivo (teachers’ own words); (3) constant comparison to collapse/expand codes; (4) theme development where recurring meanings coalesced; (5) drafting short analytic memos to define boundaries and note disconfirming evidence. Below, each inductive code is shown with what prompted it in the data (*data triggers*) and verbatim excerpts.

Inductive Code	Emergence (data triggers)	Key excerpts
Teacher-champion leadership	Teachers describing active leadership or coordination roles in rollout.	<p>“I had signed us up for it... and at our first department meeting I just explained how this was going to work and how we were going to try it throughout the year.” (STAR5)</p> <p>“It was my role initially... helping them make sure that they knew what they were doing, helping them get set up and coordinating with your end in terms of potentially issues or problems.” (STAR6)</p> <p>“I’ve been in to support teachers doing the assessments and things like that... I’ve been in three or four classes when they’ve done the assessments just to support and help a bit.” (STAR4)</p>
Set-Up	Onboarding process	<p>“It was very very seamless and easy to get it set up... definitely very user friendly.” (STAR5)</p>

		<p>"Once we got that set up, it was reading across the wrong email addresses... but again, that was an easy thing to do once I worked it out." (STAR6)</p>
Teacher Acceptability	Variation in teacher enthusiasm or digital confidence.	<p>"You probably picked the three who were I would say the non-accepting of it... anti-technology, anti-using it, blah blah blah." (STAR6)</p> <p>"We probably have one member of the department who maybe just wouldn't be as big in it and apps and things like that there." (STAR5)</p>
Diagnostic Insight	Eedi's diagnostic capacity for tracking misconceptions and informing planning.	<p>"We knew it was diagnostic in terms of identifying sort of misconceptions and then helping the students overcome them with a bit of additional support... having that extra data more consistently throughout the year was going to be a big benefit." (STAR5)</p> <p>"I said right let's take the most basic set of pie chart questions... that gave me a bit of a starting point for when I was teaching it." (STAR6)</p> <p>"I've used it as homework, reinforcement, and testing -just really quick sort of five-minute, ten-minute at the end of a lesson just to see who's got it." (STAR4)</p>
School Priorities	Eedi supporting whole-school aims	<p>"One of our staff development aims was to incorporate retrieval practice. So, it basically done the work for us." (STAR5)</p>
Differentiation and Challenge	Platform adaptability to learner levels	<p>"We are an all-ability school and we don't stream our classes... I think that was a really really good benefit for us having that aspect of it." (STAR5)</p> <p>"Teacher said... he's got set one... it was taking them a minute to do all of them and still</p>

		<p>getting [them] right... he wished there were more." (STAR6)</p> <p>"Better in sets, not so good when I've had mixed ability... that's a bit more tricky." (STAR4)</p>
Pupil Incentives and Motivation	Prizes and recognition as engagement levers	"After the first student won it they were just like, 'oh my god'... two or three of our students have won it... they really realise they have a genuine chance of winning something." (STAR5)
Embedded Practice	Different implementation patterns - homework, IT suite, cover, testing.	<p>"We aimed for one a week... either set as a homework or... booked the computer suite and maybe assigned them two or three tasks." (STAR5)</p> <p>"Homework only." (STAR6)</p> <p>"I've mainly used it as homework... and I've also used it for testing just really quick five-minute, ten-minute at the end of a lesson." (STAR4)</p>
Communicating Feedback	Weekly emails or progress reports used for reminders, accountability, or parental contact	<p>"I particularly find them useful... probably is a good reminder that if you do see that you have a week where you haven't maybe students doing a test you think, 'oh, I maybe haven't pushed that week.'" (STAR5)</p> <p>"It did give me a snapshot... it meant I didn't have to necessarily go and investigate it." (STAR6)</p> <p>"I snipped those [reports] and sent those home to parents... after the feedback from that, the parents were so on it... they were between 90 and 100% every week completion." (STAR4)</p>
Hardware and Software	Hardware or platform compatibility	"It's just a real pity that our... iPads are really really old . It doesn't actually - the Eedi doesn't work on the web browser. " (STAR5)

Time Efficiency	Platform-associated tasks and time investment	<p>"It only allows you to select one quiz at a time... it takes ages... I'd like to set two quizzes... because then it doesn't go back to the original search either." (STAR4)</p> <p>"Took a long time because I had to snip thirty different ones... if there was some way of printing a PDF... then I could just send those to admin." (STAR4)</p>
Pupil Ownership	Independence and personalised learning	<p>"They then realised that they weren't all getting the same questions and it was based off them... I thought that was good." (STAR5)</p> <p>"If you could change the order of the questions in a quiz just so it's a bit randomised... three boys just sat there and one went for every wrong answer, the other two learned from his answers." (STAR4)</p>
Technical Support	Appreciation for human contact and quick problem-solving	<p>"It helped knowing that there was kind of a face at the end of the email as well... knowing then that actually if I email Name, you can get back to me in 24 hours max." (STAR6)</p> <p>"Both you and Name have been brilliant... quick feedback every time I've been stuck." (STAR4)</p>

Table 31. Inductive coding of interview transcripts

Framework Analysis

After defining the inductive codes above, we created a deductive-inductive codebook with deductive domains aligned to survey domains (usability, access/IT, engagement, content alignment, workflow/analytics, implementation). Where there was overlap between inductive and deductive codes, these were combined and extended into major/minor codes. We then re-assigned these codes to transcripts and grouped them into major themes. This led to the development of 5 major themes that were explored in relation to survey and RCT results.

Leadership, Culture, and Support

Teachers consistently described the importance of internal champions and responsive external support in sustaining implementation. This suggests that implementation success may be related to local leadership and support culture. Several teachers acted as internal “champions”, initiating the project and supporting colleagues.

“I had signed us up for it... and at our first department meeting I just explained how this was going to work and how we were going to try it throughout the year.” (STAR5)

“It was my role initially... helping them make sure that they knew what they were doing, helping them get set up and coordinating with your end in terms of potentially issues or problems.” (STAR6)

“I’ve been in to support teachers doing the assessments... I’ve been in three or four classes when they’ve done the assessments just to support and help a bit.” (STAR4)

Teachers also valued the personal and rapid support provided by the developer team:

“It helped knowing that there was kind of a face at the end of the email as well... knowing then that actually if I email you, you can get back to me in 24 hours max.” (STAR6)

“Brilliant... quick feedback every time I’ve been stuck.” (STAR4)

“I just like the fact that you can actually talk to somebody... it makes such a difference because otherwise teachers just switch off.” (STAR6)

These comments reflect the survey results, where nearly half of respondents found setup somewhat or very straightforward. While the RCT revealed variable impacts between schools, these accounts explain some of that variation: where internal champions and responsive support existed, implementation fidelity may have been higher.

Embedded Pedagogy: Diagnostics, Retrieval, and Differentiation

Teachers repeatedly emphasised Eedi’s diagnostic value and alignment with school priorities.

“We knew it was diagnostic in terms of identifying misconceptions and then helping the students overcome them... having that extra data more consistently throughout the year was going to be a big benefit.” (STAR5)

“One of our staff-development aims was to incorporate retrieval practice, so it basically done the work for us.” (STAR5)

"I used it for recap and prior understanding; it gave me a bit of a starting point before I taught pie charts." (STAR6)

"We are an all-ability school and we don't stream our classes... having that aspect of it [automatic adaptation] was really really good." (STAR5)

Yet differentiation sometimes remained challenging, with top sets appearing to complete tasks quickly, whilst lower-ability students found tasks more difficult which was identified as detrimental to their progress and confidence.

"Better in sets, not so good when I've had mixed ability... that's a bit more tricky." (STAR4)

"Teacher said his top set whizzed through it - it was taking them a minute to do all of them and still getting them right - he wished there were more." (STAR6)

"Some of the lower groups struggled when it ramped up too quickly... they'd get stuck and then lose confidence." (STAR6)

Survey data show that 13 of the 19 teachers who used Eedi in class (68%) used its diagnostic questions, while 61% of teachers flagged curriculum alignment as a barrier. This closely mirrors the mixed-ability tension described here. The RCT's small but positive average effect ($d \approx 0.10$, not statistically significant) with wide between-school variation likely reflects settings where diagnostic alignment and retrieval fit were strong versus those where pitch mismatch limited progress. This may explain why Eedi was not consistently used in class.

A further challenge to embedding Eedi within classroom practice was teacher willingness and confidence in using digital platforms. This emerged as a potentially significant moderating factor in the intensity and type of Eedi use. While overall attitudes toward the platform were positive, several teachers described colleagues who were hesitant or resistant to classroom technology.

"You probably picked the three who were I would say the non-accepting of it... anti-technology, anti-using it, blah blah blah." (STAR6)

"I've been in to support teachers doing the assessments... because people weren't always up with it." (STAR4)

This could explain why many survey respondents opted to assign quizzes as homework rather than integrate them into lessons, and in turn, provides a plausible explanation for the relatively low average dosage observed in the RCT and for the variation in effects across schools. Although 82% of intervention teachers attended training and 49% found set-up somewhat or very straightforward, only around a third reported

using Eedi during lessons (34% of intervention-only and 41% of dual-group teachers), and most platform use took the form of homework. Respondents frequently mentioned “preferring to keep tech homework-based” or avoiding “logging students in mid-lesson.” Teachers who reported lower confidence were also those least likely to view in-class diagnostics as valuable, instead using Eedi for reinforcement tasks outside class time.

The RCT dosage data showed that, across the intervention arm, average pupil completion of quizzes fell below the intended target of one per week, with wide variation between schools. This aligns closely with the qualitative accounts of partial implementation. Where teachers preferred to assign homework, engagement was constrained by home-access issues, which the survey identified as a major barrier for 71 % of teachers. The prior problems with differentiation and accessibility however means when assigned as stand-alone homework, lower-attaining pupils may not have been as engaged or supported when encountering difficulties.

Engagement and Motivation Loops

Teachers identified a number of approaches that were successful in driving pupil engagement, ranging from Eedi’s progress emails and raffles alongside teacher-produced parental updates.

“After the first student won it... two or three of our students have won it now; they really realise they have a genuine chance of winning something.” (STAR5)

“I snipped those reports and sent them home to parents with a little blurby email... after the feedback from that the parents were so on it that they really wanted the pupils doing it.” (STAR4)

“That class were between 90 and 100 percent every week completion when I was really on it... just getting those parents involved really helped.” (STAR4)

“I particularly find the weekly emails useful; it’s a good reminder... keeping you to account if you’ve had a quiet week.” (STAR5)

“They then realised they weren’t all getting the same questions and it was based off them; that made them much keener to complete it.” (STAR5)

“Once we put the leaderboard on the board on a Friday, they actually asked for more homework.” (STAR6)

While 56 % of teachers reported homework completion above 75 %, motivation and home-access barriers remained high. Interview evidence shows that extrinsic rewards,

parental feedback, and visible progress tracking countered these challenges. These mechanisms plausibly underlie the directionally positive RCT effect despite non-significance: engagement loops raised completion and hence potential learning time.

Usability and Access

Most teachers found the system intuitive but were sometimes limited by hardware and connectivity.

"It was very very seamless and easy to get it set up... definitely very user-friendly." (STAR5)

"Once we got that set up, it was reading across the wrong email addresses, but again that was an easy thing to do once I worked it out." (STAR6)

"It's just a real pity that our iPads are really really old; Eedi doesn't work on the web browser." (STAR5)

"I love the new 'check' feature down the side... only when you use it you realise how useful it is." (STAR4)

"Sometimes the connection drops in the IT suite and you've lost half of them before they log back in." (STAR6)

The survey similarly found that setup was straightforward for most users, yet nearly half reported school-IT barriers. These structural constraints may explain uneven dosage: classes with reliable devices might have been more likely to use the tool routinely, while those with aging or blocked hardware could not, which may have contributed to between-school variability in RCT outcomes.

Workflow and Analytics

Teachers praised the clarity of Eedi's progress data but wanted greater efficiency in assigning and reporting tasks.

"It did give me a snapshot; it meant I didn't have to necessarily go and investigate" (STAR6)

"The reports where you've got the pupil's name and the bar for effort and completion, those are brilliant." (STAR4)

"It only allows you to select one quiz at a time... it takes ages... I'd like to set two quizzes." (STAR4)

"If you could change the order of the questions in a quiz just so it's a bit randomised... three boys sat together and one went for every wrong answer, the other two learned from him." (STAR4)

"It would be great if you could just tick multiple quizzes and set them in one go, it would save so much time." (STAR6)

Most teachers in the survey reported neutral or reduced workload, suggesting time savings through automated marking. However, open-text comments suggest some workflow concerns.

Across interviews, survey, and trial data, the evidence converges on several key mechanisms: strong teacher leadership, clear diagnostic alignment, motivational feedback loops, and rapid developer support were amplifiers of impact, whereas access issues, mixed-ability misalignment, and workflow inefficiencies may have constrained impact. These mechanisms could explain why the RCT demonstrated directionally positive but variable outcomes: benefits emerged where diagnostic use, motivation, and support combined, but were dampened where access and differentiation barriers persisted.

Convergence Coding Matrix

The matrix below synthesises RCT, survey, and interview evidence in a joint display organised around five themes: Leadership, Culture, and Support, Embedded Pedagogy: Diagnostics, Retrieval, and Differentiation, Engagement and Motivation Loops, Usability and Access, Workflow and Analytics. For each construct we indicate whether strands Converge (C), Complement (Comp), Diverge (D), or are Silent (S). Analysis across the RCT, survey, and interview strands points to a coherent account of why outcomes varied between schools. The data converge on three enabling conditions that, when present together, may lead to higher teacher uptake, stronger pupil engagement, and improved dosage of use. First, leadership and support were critical: departments with a clear "Eedi champion" who coordinated setup and modelled described smoother implementation, especially when backed by prompt, personal support from the Eedi team. Second, pedagogical fit mattered: teachers who wove Eedi's diagnostic and retrieval tools into their existing schemes of work would have increased the likelihood of pupils using Eedi consistently and effectively. Third, engagement loops such as parental reporting, leaderboards, and small rewards can sustain motivation and drive completion rates. Where these factors co-occurred, it is likely that usage was higher, and implementation more consistent.

Several contextual factors limited how often and how effectively Eedi was used. Access barriers, including outdated school devices, browser issues, and unreliable connectivity, pushed activity out of lessons and onto pupils' home devices, where engagement depended on variable levels of technology and parental support. Teacher digital reluctance also played a role, with some staff preferring to set Eedi only as homework to avoid technical disruption in class, reducing opportunities for in-the-moment feedback and widening gaps for less supported students. Finally, workflow friction, such as having to assign quizzes individually, limited question randomisation, and short task lengths, may have discouraged use even though most teachers felt overall workload was manageable.

Major Theme 1: Leadership, Culture and Support				
<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	ITT small, +ve, non-sig; between-school variation (some +ve, some neutral).	82 % attended training; 49 % rated setup somewhat or very straightforward; only 16 % found it challenging.	Leadership and local champions, variation in buy-in, human support valued.	Comp (small average + heterogeneity fit a possible pattern of stronger outcomes where leadership/support were robust).
Internal Capacity + Champions	No direct RCT metric; higher-fidelity schools showed stronger local effects.	No quantitative item on departmental leads. Free-text mentions leadership or peer support as facilitators.	Specific teachers coordinated setup, trained peers, modelled in-lesson use.	Comp (some free-text survey suggestion via free-text; interviews extend explanation).
Developer support responsiveness	Not measured directly; plausibly affects sustained use.	85 % reported technical support was timely and effective.	Very positive feature of interview responses	C (strong cross-strand agreement).

Cultural buy-in, digital confidence	School-level variation suggests local culture may moderate impact.	47 % cited staff digital-confidence barriers.	Interviews referred to digital-confidence barriers.	C (strong cross-strand agreement).
-------------------------------------	--	---	---	---

Major Theme 2: Embedded Pedagogy – Diagnostics, Retrieval, and Differentiation

<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	Per-protocol ES direction positive; school-level effects mixed.	68 % used diagnostic quizzes; 61 % cited curriculum alignment/difficulty as barriers.	Positive perspectives on utility of diagnostic tools.	Comp (quantitative and qualitative strands describe same mechanism).
Differentiation & set-fit	No subgroup significant; descriptive data suggest stronger trends in high-attaining groups	Free-text comments highlighted some tasks were too difficult for bottom-set groups.	Reference to more challenges with mixed and lower ability sets, top-sets completing challenges quickly and required more work.	C (alignment across strands).
Retrieval & curriculum alignment	Positive trends with frequent use.	Responses linked Eedi to retrieval practice aims.	Responses linked Eedi to retrieval practice aims.	C (alignment across strands).

Major Theme 3: Engagement and Motivation Loops

<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	Directionally +ve ITT; per-protocol ES suggests engagement ↔ impact.	56 % ≥ 75 % homework completion; 72 % cited motivation, 71 % home access as barriers.	Number of features identified that were felt to increase pupil engagement and buy-in.	Comp (survey quantifies engagement issues; interviews show motivating factors).

Parents & home	Not measured.	Generally high rates of homework completion but noted concerns about pupil access to devices when at home.	Some teachers used Eedi reporting to share feedback with parents and felt this increased engagement.	Comp (survey quantifies engagement issues; interviews show motivating factors).
Incentives	Not measured.	Free-text responses refer to awards and prizes increasing pupil engagement.	Number of features identified that were felt to increase pupil engagement and buy-in.	Comp (survey quantifies engagement issues; interviews show motivating factors).

Major Theme 4: Usability and Access				
<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	Small +ve average.	Setup easy (49 % "somewhat" or "very straightforward"); home access 71 %, school IT 49 % barriers.	Generally very positive experiences of using Eedi.	Comp (survey quantifies engagement issues).
Hardware & connectivity	Not directly measured but could drive dosage variation.	Top-rated barrier category.	Little accessibility issues noted in interviews.	Comp (survey quantifies engagement issues).
Ease of setup & interface	Not measured.	49 % found setup somewhat or very straightforward; positive free-text on simplicity.	Some set-up issues noted but easily and quickly resolved.	C (alignment across strands).
Digital reluctance / in-class avoidance	No direct metric; indirectly captured by low lesson-use frequency.	47 % acknowledged staff confidence issues; most teachers used Eedi mainly for homework (34-41% reported in-lesson use).	Identified Eedi was largely used as a homework tool, also highlighted some staff reluctance.	C (alignment across strands).

Major Theme 5: Workflow and Analytics
--

<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	Not measured.	Time savings for workload noted.	Teachers report not having to spend as long investigating pupil understanding.	C (alignment across strands).
Assignment efficiency	Not measured.	Teachers requested bulk-set assignment and multi-select features	Some reference to requests for greater functionality.	C (alignment across strands).
Reporting & analytics usability	Not measured.	Monitoring rated straightforward and time-saving.	Generally positive responses to reporting and analytics although noted it can be time consuming to share	C (alignment across strands).

Discussion

This evaluation provides a detailed account of how Eedi was implemented across participating schools and the outcomes that followed. The findings present a nuanced picture. On average, the intervention produced a small, positive but statistically non-significant improvement in pupil attainment, with stronger effects observed in schools and classes where engagement and compliance were higher. Teachers generally viewed Eedi as accessible and manageable, reporting neutral or reduced workload and valuing its diagnostic quizzes, automated feedback, and homework-monitoring tools. However, implementation varied widely. Consistent use was often limited by student motivation, digital access at home, and differences between Eedi's content and school curricula. Some teachers, particularly those less confident with technology, preferred to use Eedi for homework rather than in lessons. This discussion explores these patterns in greater depth, drawing on triangulated evidence from the RCT, teacher surveys, and interviews to examine the conditions under which Eedi achieved its strongest impacts and the practical factors that constrained its effectiveness.

Intervention Effectiveness and Effect Size in Context

The RCT found that Eedi produced a small, positive but statistically non-significant improvement in Year 7 mathematics attainment. After adjusting for baseline scores, pupils in the treatment group scored on average 4.99 points higher than those in the control group (effect size ≈ 0.10 ; 95% CI = -0.08 to 0.28), with a confidence interval spanning zero. This suggests that while the direction of effect was favourable, the overall impact was modest and uncertain. These findings are consistent with the broader evidence base on educational technology, where interventions typically yield small positive effects when implemented in authentic classroom conditions. The estimated effect size for Eedi is comparable to that reported in other rigorous EdTech trials and falls within the expected range for school-based interventions delivered at scale.

Further analysis revealed that outcomes varied considerably between schools. While several schools showed small positive differences favouring the treatment group,

others displayed neutral or slightly negative effects. This variation is mirrored in the Implementation and Process Evaluation, which highlighted differences in teacher engagement, technical access, and integration within existing schemes of work. In particular, the Complier Average Causal Effect (CACE) analysis indicated stronger outcomes among students and teachers with higher engagement levels, suggesting that the intervention was more effective when implemented with fidelity and regular use. At the highest compliance threshold (≥ 200 diagnostic questions), the effect size increased to approximately 0.24 (95% CI 0.03 to 0.44), the threshold at which the confidence interval excluded zero; these complier estimates are exploratory. Among Pupil Premium pupils at this threshold, the effect size was substantially larger ($g \approx 0.57$; 95% CI 0.32 to 0.82). These patterns point to the importance of dosage and context. Although the intervention ran across a full academic year, many teachers reported using Eedi mainly for homework rather than during lessons, and overall pupil completion rates were below the intended weekly target. Implementation data also revealed that barriers such as limited home internet access, misalignment with school curricula, and variable digital confidence among teachers constrained consistent use. Findings suggest that the modest average effect observed in the RCT reflects differences in implementation and access rather than a lack of underlying efficacy.

Leadership, Culture and Support

Implementation success appeared closely linked to school leadership, departmental culture, and teacher confidence. Although overall attitudes towards Eedi were positive, levels of teacher engagement varied considerably across schools. The Implementation and Process Evaluation (IPE) indicated that 82% of intervention teachers attended training and almost half found set-up somewhat or very straightforward, yet survey and interview data also revealed uneven buy-in within departments. Several teachers described colleagues who were hesitant or resistant to using technology in class, a finding consistent with interview accounts and survey free-text responses suggesting that limited digital confidence and staff preference for traditional approaches constrained in-lesson use. In many schools, this variation in confidence shaped how Eedi was implemented. Teachers with higher digital confidence and departmental support were more likely to integrate Eedi into lessons, using its diagnostic quizzes and data to inform teaching. Others, particularly those less comfortable with classroom technology, tended to restrict use to homework. While this still provided opportunities for additional practice, it reduced exposure to Eedi's diagnostic and feedback functions that were intended to support formative assessment. Interviews highlighted the importance of internal "champions" who coordinated setup, encouraged colleagues, and liaised with the Eedi team. This pattern helps explain the variability in outcomes

observed in the RCT. The overall effect was small and not statistically significant, but CACE analysis showed larger positive effects among classes with higher engagement. Together, these findings suggest that leadership, supportive culture, and teacher confidence were key enabling conditions for achieving the potential benefits of Eedi.

Embedded Pedagogy

Teacher interviews and survey responses indicated that several core features of Eedi, particularly its diagnostic questioning and retrieval practice functions, aligned well with teachers' instructional aims. Many teachers valued how the quizzes provided quick insights into pupils' understanding and misconceptions, supporting formative assessment and lesson planning. This alignment with diagnostic and retrieval-focused pedagogy reinforces Eedi's intended theory of change and helps explain why teachers with higher engagement levels reported positive experiences. However, while Eedi was seen as a useful tool for identifying learning needs and promoting retrieval practice, it was not consistently embedded within daily teaching across schools. Most teachers reported using the platform primarily for homework rather than as an integrated classroom resource, reflecting variable implementation fidelity across sites.

Challenges around differentiation and curriculum alignment were the most frequently cited pedagogical limitations. Teachers in mixed-ability classes reported that quiz difficulty levels were not always well calibrated: high-attaining pupils often found the quizzes too easy, while lower-attaining pupils struggled when question difficulty increased too quickly. This lack of scaffolding and stretch limited perceived accessibility for both groups. These findings were echoed in the IPE survey, where 61 % of teachers identified misalignment between Eedi's quiz sequencing and their own scheme of work as a barrier, and 22% considered this a significant issue. Several teachers noted that Eedi quizzes occasionally covered content not yet taught or used unfamiliar methods, which caused confusion for some pupils and reduced confidence.

These alignment issues likely influenced how frequently and effectively Eedi was used. Teachers who perceived better curricular fit were more likely to integrate Eedi into their planning and formative assessment routines, while others restricted use to optional or supplementary homework tasks. This partial integration helps explain the moderate average dosage observed in the RCT and the small, non-significant effect on attainment. Moreover, where Eedi was mainly used for homework, differences in pupil access to devices and internet connectivity meant that engagement may have varied, particularly for lower-attaining pupils. Taken together, the evidence suggests that Eedi's diagnostic and retrieval features were pedagogically sound and well-received,

but their impact depended on how effectively the platform was aligned with local curricula and differentiated for diverse learners.

Engagement and Accessibility

IPE findings suggest that Eedi was broadly user-friendly, accessible, and well-received by teachers. The majority of teachers (82%) attended onboarding training, and nearly half (49%) described setup as somewhat or very straightforward, while only a small minority (16%) found it challenging. These results indicate that most teachers were able to integrate Eedi with minimal technical support.

Perceptions of workload were largely positive. Almost half of intervention-only teachers (45%) and half of dual-group intervention teachers (50%) reported no change in overall workload, while a further 34-44% perceived a decrease. One in ten intervention-only teachers and just under a quarter of dual-group intervention teachers reported an increase (Table 25). Teachers attributed these efficiencies mainly to automatic marking, simplified tracking of homework completion, and reduced time spent preparing resources or reviewing common errors. Whilst intervention teachers reported spending an average of 49-56 minutes per week on homework administration, compared with 22-34 minutes among control teachers, they suggest that any additional time spent setting up Eedi quizzes was offset by reductions in marking and feedback effort. Teachers frequently commented that automated marking and instant feedback streamlined the assessment–feedback loop, making routine homework management more efficient.

The only group who reported potential workload increases were teachers managing both intervention and control classes. This was likely due to the additional administrative complexity of running two parallel systems rather than the platform itself. Even within this group, half reported no change and over a quarter still reported reduced workload.

Despite these positives, teachers identified several workflow limitations that constrained usability. The most common issue was the inability to assign multiple quizzes at once, which some described as time-consuming. Others noted that the default five-question format felt too short, particularly for high-attaining classes, and that lack of randomisation limited opportunities for independent practice. These were framed as minor frustrations rather than barriers to adoption: teachers continued using the platform but viewed these adjustments as areas for improvement.

Conclusion

This evaluation provides a balanced and evidence-based picture of how Eedi functioned in school settings. The RCT found a small, positive but statistically non-significant improvement in mathematics attainment for Year 7 pupils using Eedi, with an average adjusted difference of around five points (effect size ≈ 0.10 , 95 % CI = -0.08 to 0.28). While modest, this effect is consistent with those reported for other large-scale, low-cost educational technology interventions. Importantly, the trial found no evidence that the intervention widened the attainment gap between Pupil Premium and non-Pupil Premium students. Point estimates were positive for both groups and somewhat larger for Pupil Premium pupils; this suggests Eedi can be used equitably across different pupil backgrounds when access is supported.

Teacher feedback provides further insight into these outcomes. Most teachers viewed Eedi as user-friendly and workload-neutral, with many reporting time savings through automated marking and feedback. These efficiencies, combined with positive attitudes toward Eedi's diagnostic and retrieval features, indicate that the platform can streamline assessment and progress monitoring without adding administrative burden. Such indirect benefits are valuable in a context where workload and teacher wellbeing remain national priorities.

At the same time, the evaluation highlights that Eedi's impact depended heavily on local implementation conditions. Three enabling factors emerged as central to effective use: strong departmental leadership and teacher buy-in, good alignment between Eedi's content and school curricula, and active strategies to sustain pupil engagement. Where these conditions were met, usage tended to be higher and benefits clearer. Conversely, limited digital access, difficulty differentiating for mixed-ability classes, and minor workflow inefficiencies were identified as recurrent barriers that restricted consistent use in some schools. These challenges often led to partial or homework-only implementation, which likely diluted overall impact.

In summary, the evaluation demonstrates that Eedi has clear potential as a practical, workload-efficient tool for supporting maths learning. The mixed but generally positive outcomes observed reflect variation in access, engagement, and curriculum alignment rather than inherent shortcomings in the platform's design. When implemented under supportive conditions with adequate technology infrastructure, leadership support, and curricular fit, Eedi can be effectively integrated into classroom practice, offering both pedagogical and operational benefits for teachers and pupils alike.

References

- Akın, A. (2022). The effectiveness of web-based Mathematics instruction (WBMI) on K-16 students' mathematics learning: a meta-analytic research. *Education and Information Technologies*, 27(6), 8015-8040. <https://doi.org/10.1007/s10639-022-10931-x>
- Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). Inequalities in children's experiences of home learning during the COVID-19 lockdown in England. *Fiscal Studies*, 41(3), 653-683. <https://doi.org/10.1111/1475-5890.12240>
- Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using Zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods*, 18. <https://doi.org/10.1177/1609406919874596>
- Benavides-Varela, S., Callegher, C. Z., Fagiolini, B., Leo, I., Altoè, G., & Lucangeli, D. (2020). Effectiveness of digital-based interventions for children with mathematical learning difficulties: A meta-analysis. *Computers & Education*, 157, 103953. <https://doi.org/10.1016/j.compedu.2020.103953>
- Betts, J. R., Hahn, Y., & Zau, A. C. (2017). Can testing improve student learning? An evaluation of the mathematics diagnostic testing project. *Journal of Urban Economics*, 100, 54-64.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Di Pietro, G., & Castaño Muñoz, J. (2025). A meta-analysis on the effect of technology on the achievement of less advantaged students. *Computers & Education*, 226, 105197.
- Durkin, K., & Rittle-Johnson, B. (2015). Diagnosing misconceptions: Revealing changing decimal fraction knowledge. *Learning and Instruction*, 37, 21-29. <https://doi.org/10.1016/j.learninstruc.2014.08.003>
- Escueta, M., Nickow, A., Oreopoulos, P., & Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4), 897-996. <https://doi.org/10.1257/jel.20191507>

Farmer, T., Robinson, K., Elliott, S. J., & Eyles, J. (2006). Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research*, 16(3), 377-394. <https://doi.org/10.1177/1049732305285708>

Feng, M., Heffernan, N., Collins, K., Heffernan, C., & Murphy, R. F. (2023, June). Implementing and evaluating ASSISTments online math homework support at large scale over two years: Findings and lessons learned. *International Conference on Artificial Intelligence in Education* (pp. 28-40). Cham: Springer Nature Switzerland.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13, 117. <https://doi.org/10.1186/1471-2288-13-117>

Galimova, E. G., Oborsky, A. Y., Khvatova, M. A., Astakhov, D. V., Orlova, E. V., & Andryushchenko, I. S. (2024, November). Mapping the interconnections: a systematic review and network analysis of factors influencing teachers' technology acceptance. *Frontiers in Education*, Vol. 9, p. 1436724. Frontiers Media SA.

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine*, 13(6), 554-561. <https://doi.org/10.1370/afm.1865>

Harrison, W., Dobson, E.S., Higgins, S., Uwimpuhwe, G. & Khowaja, R. (2025). A study to evaluate the effectiveness of Eedi on raising attainment in mathematics at KS3 (Year 7). Durham, England: WhatWorked Education. https://cdn.prod.website-files.com/68b00fd42843f377c633738c/68fba1cac6352a100f8aa11e_Eedi_RCT_Impact_Report_Final.pdf.pdf

Hettinger, K., Lazarides, R., & Schiefele, U. (2022). Motivational climate in mathematics classrooms: Teacher self-efficacy for student engagement, student- and teacher-reported emotional support and student interest. *ZDM - Mathematics Education*, 55(2), 413-426. <https://doi.org/10.1007/s11858-022-01430-x>

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3-20.

<https://doi.org/10.1177/1525822X05282260>

J-PAL Evidence Review. 2019. "Will Technology Transform Education for the Better?" Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.

<https://www.povertyactionlab.org/publication/will-technology-transform-education-better>

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, 2, 100017. <https://doi.org/10.1016/j.caeai.2021.100017>

Kucirkova, N., Schewe, O., Campbell, J., Cermakova, A. L., & Pitchford, N. (2025). Developing evidence indicators for evaluating K12 EdTech: Towards a consensus on educational impact. *Humanities & Social Sciences Communications*, 12, Article 947. <https://doi.org/10.1057/s41599-025-05330-9>

Létourneau, A., et al. (2025). A systematic review of AI-driven intelligent tutoring systems in K-12 education. *npj Science of Learning*, 10, Article 24.

<https://doi.org/10.1038/s41539-025-00320-7>

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in Education*. London: Pearson.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901-918. <https://doi.org/10.1037/a0037123>

Maldonado, J. E., & De Witte, K. (2022). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*, 48(1), 49–94.

<https://doi.org/10.1002/berj.3754>

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. <https://doi.org/10.1177/1049732315617444>

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47. <https://doi.org/10.1177/016146811311500303>

Meylani, R. (2025). AI-Powered Assessments in Mathematics Education: A Systematic Review of Contemporary Research Literature. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (66), 3642-3674.

Nickow, A., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. *American Educational Research Journal*, 61(1), 74-107.
<https://doi.org/10.3102/00028312231208687>

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84.
<https://doi.org/10.3102/0034654307313793>

Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2021). Effective programs for elementary mathematics: A best-evidence synthesis. *AERA Open*, 7(1), 1–17.
<https://doi.org/10.1177/2332858420986214>

Putri, A. D., & Juandi, D. (2025). Blended Learning and Math Achievement: A Meta-Analytic Review Highlighting the Effectiveness and Heterogeneity. *Electronic Journal of e-Learning*, 23(1), 113-128.

Ran, H., Kim, N. J., & Secada, W. G. (2022). A meta-analysis on the effects of technology's functions and roles on students' mathematics achievement in K-12 classrooms. *Journal of computer assisted learning*, 38(1), 258-284.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA open*, 2(4), 2332858416673968.

Samritin, S., Susanto, A., Manaf, A., & Hukom, J. (2023). A meta-analysis study of the effect of the blended learning model on students' mathematics learning achievement. *Jurnal Elemen*, 9(1), 15-30.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The Technology Acceptance Model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13–35.
<https://doi.org/10.1016/j.compedu.2018.09.009>

Setiawan, A. A., Muhtadi, A., & Hukom, J. (2022). Blended Learning and Student Mathematics Ability in Indonesia: A Meta-Analysis Study. *International Journal of Instruction*, 15(2), 905-916.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–407.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

van der Kleij, F. M. (2019). Clarifying the formative feedback process with the Assessment for Learning Questionnaire. *Studies in Educational Evaluation*, 62, 1–10. <https://doi.org/10.1016/j.stueduc.2019.04.001>

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>