

Technical Report

The Impact of Sustained Use of Eedi on KS3 Mathematics Attainment

Evidence from a Two-Year Cluster
Randomised Controlled Trial (Years 7-8)

Table of Contents

Table of Contents.....	2
Executive Summary.....	3
Background.....	5
Methods.....	12
Impact Results: Summary.....	20
Impact Results: Technical Analysis.....	21
IPE Results: Summary.....	30
IPE Survey Results: Technical Analysis.....	32
IPE Interview Results.....	48
Discussion.....	56
Conclusion.....	59
References.....	61

Executive Summary

This evaluation examined the impact of *Eedi*, a digital homework and assessment platform designed to support Key Stage 3 mathematics through diagnostic questioning, retrieval practice, and adaptive feedback. The study used a cluster randomised controlled trial (RCT) with schools randomly assigned to either use *Eedi* (intervention) or continue with usual practice (control). The trial ran over two academic years (2023-25) and was complemented by an Implementation and Process Evaluation (IPE) combining teacher surveys, platform data, and interviews.

Participants

The first year of the study involved 12 intervention and 8 control schools. For the extended two-year phase, all participating schools were invited to continue, and nine schools chose to do so. Accordingly, the number of schools providing data at 18- and 24-month follow-up was nine.

Key Findings

- **Positive overall effects:**

Across the 24-month period, average treatment effects were positive at all time points and increased from small at six months ($d = 0.06$) to large by 18 months ($d = 0.46$), before stabilising at 24 months ($d = 0.30$). Estimates were statistically significant at 18 and 24 months. When averaging across all time-points, the intervention effect ($d = 0.25$) was statistically significant.

- **Consistency matters:**

The CACE (Complier Average Causal Effect) analyses showed that pupils who engaged more consistently achieved larger and almost always statistically significant gains, particularly at 18 months. This confirms that *Eedi's* impact is implementation-dependent, with meaningful benefits emerging when the platform is used as intended.

- **Lower-attaining pupils show promise:**

Sensitivity analyses found that pupils starting with lower baseline attainment made important gains, peaking at 18 months ($d = 0.415$). This suggests that *Eedi* may be particularly valuable for learners who need greater scaffolding and feedback. Additionally, the effect of intervention in these students was significant at 12, 18 and 24 months.

- **Teacher experience and workload benefits:**

Teachers reported reduced marking loads, more efficient progress monitoring,

and actionable diagnostic insights. They also highlighted curriculum mapping gaps, device access constraints, and variation in homework culture as factors limiting consistent usage.

Implications

The findings demonstrate that Eedi is a usable, teacher-friendly platform with strong potential to improve mathematics learning when used consistently. Encouraging regular weekly retrieval practice and strengthening scaffolds for lower-attaining pupils could enhance both uptake and effect size.

The impact of the intervention along with the significant adherence effects highlight that measurable learning gains are achievable under realistic classroom conditions. The evaluation supports continued investment in implementation support, teacher training, and iterative refinement, ensuring that Eedi's strongest mechanisms - retrieval, feedback, and diagnostic precision - translate into sustained, scalable impact.

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). 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., 2023), 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 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 8 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). This trial is a follow-on to a previous RCT in Year 7 (2023–24) and therefore provides a unique opportunity to assess whether usage and impact is sustained after two years of implementation. 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. By combining impact estimates with implementation and process evidence, the study is positioned to test not only whether Eedi improves KS3 mathematics attainment, but also whether sustained use and fidelity help explain when and for whom benefits emerge.

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 to two years of using Eedi on students' mathematics attainment, as measured by the Year 8 NWEA MAP maths test?
- **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. Whole schools taking part in the study were randomly assigned to one of two groups:

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

By randomly assigning schools on a 1:1 basis, any differences in pupil outcomes can be attributed to the use of *Eedi*, rather than to pre-existing differences between schools or students. The initial study period of 12 months was extended to 24 months, with participating schools invited to remain involved for a second year of implementation.

Randomisation

Schools were randomly assigned on a 1:1 basis to either the intervention group (with access to *Eedi* for their Year 7 cohort during the 2023-24 academic year and Year 8 cohort for the 2024-2025 academic year) or the control group (who continued with their usual maths provision). Randomisation was carried out by the *WhatWorked Education* research team, who remained blind to the identity of participating schools throughout the process.

To meet data-sharing requirements, all schools were anonymised and allocated a unique identification code. Using these IDs, *Eedi* securely linked school-level data from the Department for Education (DfE) performance tables and the *WhatWorked Education* School Lookup database. The dataset included contextual information such as each school's Progress 8 and Attainment 8 scores, the percentage of pupils eligible for Pupil Premium (PP), the proportion of pupils with English as an additional language (EAL), and school type.

Participants

The study initially aimed to recruit 20 schools for the evaluation. It was estimated that each school would have around 200 Year 7 pupils, giving an expected total sample of approximately 4,000 pupils, with 2,000 in the *Eedi* (intervention) group and 2,000 in the control group.

In practice, 33 schools agreed to take part. One school was excluded prior to randomisation because it was an alternative provision setting with a very high proportion of Pupil Premium students (97%) and low Progress 8 and Attainment 8 scores. This resulted in a baseline sample of 32 schools, which were then randomly allocated to the intervention (n = 16) or control (n = 16) groups.

Subsequent implementation challenges with NWEA testing reduced the active sample to 23 schools. A further three schools withdrew during the study period, leaving 12 intervention and 8 control schools. Within the intervention group, one school was unable to complete the midpoint (6-month) assessments due to staff illness but successfully completed the final (12-month) assessments.

For the extended two-year phase, all participating schools were invited to continue, and nine schools chose to do so. Accordingly, the number of schools providing data at the 18- and 24-month follow-ups was nine.

Outcome Measures

Pupil progress was measured using the NWEA MAP Growth assessment, which provides a reliable way to track learning over time. The evaluation followed a longitudinal cohort design, establishing a baseline in Year 7 and tracking the same group of pupils across five key assessment points:

1. Baseline (start of Year 7)
2. Mid-year (6 months)
3. End of Year 7 (12 months)
4. Mid-Year 8 (18 months)
5. End of Year 8 (24 months)

Results are reported using RIT scores (Rasch unit), a stable, equal-interval scale used by MAP Growth. Because RIT scores are grade-independent, a score of 220 represents the same level of achievement whether it is earned in Year 7 or Year 8. This feature makes the RIT scale particularly valuable for longitudinal analysis, allowing meaningful comparison of student growth both within and across school years. By following the same pupil cohort throughout the 24-month study, the design avoids cohort effects and provides a robust framework for understanding both individual and group progress over time.

Intervention

Eedi is a digital learning platform that offers targeted quizzes designed to support mathematics teaching at Key Stage 3 (KS3). The online quizzes provide pupils with immediate feedback, short video explanations, and the option to access on-demand one-to-one support from a qualified teacher through a chat-based feature. Designed to complement classroom teaching, *Eedi* can be used both during lessons and as homework. The platform also provides teachers with diagnostic insights into common misconceptions, a presentation mode for sharing content with the class, and reporting tools to track pupil progress over time. Participating schools were given flexibility in how they implemented *Eedi*, allowing the study to evaluate its impact under realistic classroom conditions. While this approach enhanced ecological validity, reflecting how schools are likely to use *Eedi* in practice, it also introduced variation in implementation, which may have made it more challenging to identify a single, clear overall effect.

Eedi Usage Data

Alongside measuring overall impact, the study also examined how *Eedi* was used in practice. Detailed usage data were collected to capture both teacher and pupil engagement with the intervention.

At the teacher level, usage records provided insights into implementation fidelity, for example, how often and how consistently quizzes were assigned through the platform. These measures help to show how closely classroom use of *Eedi* aligned with its intended design, and the extent to which teachers integrated it into their regular teaching routines.

At the pupil level, data were collected on quiz completion rates, time spent on activities, and interactions with tutors via *Eedi*'s chat-based support feature. These tutor interaction data reveal how frequently students sought help, the types of questions they asked, and the balance between independent practice and guided support.

Together, these data provide a rich picture of real-world use, showing how teachers and students engaged with *Eedi* across different school contexts. By combining evidence on teacher implementation, student participation, and tutor support, the usage analysis offers valuable context for understanding not only whether *Eedi* had an impact, but also how it was experienced and embedded in practice.

Primary Analysis

All statistical analyses were conducted using R Studio. The analysis plan was designed to evaluate the long-term impact of *Eedi* through a cluster randomised controlled trial, in which whole schools were randomly assigned to either the intervention or control group. The primary objective was to estimate the effect of *Eedi* on pupils' academic achievement, as measured by NWEA MAP Growth (RIT) scores over a 24-month period. Secondary objectives included exploring whether the impact differed by Pupil Premium status and examining how the effect evolved across multiple time points. The study followed a longitudinal design with assessments at five points: baseline (pre-test), 6 months (mid-test), 12 months (post-test), 18 months (continuation mid-test) and 24 months (continuation end-test). The primary outcome measure was the NWEA TestRITScore, a continuous indicator of student performance. Because students were nested within schools, creating a hierarchical data structure, the analysis used multilevel (mixed-effects) modelling to account for clustering effects. Baseline covariates included pre-test scores and Pupil Premium (PPFSMIndicator) status, which was used as a key variable in subgroup analyses to examine differential effects across pupil groups.

Primary Statistical Modelling

The primary analysis used a three-level hierarchical linear mixed-effects model to reflect the nested structure of the data, with repeated measurements over time within pupils, and pupils within schools. The model included fixed effects for:

- the between-school and within-school components of baseline (pre-test) scores (using school-level centering),
- the treatment condition (intervention or control),
- the time points of measurement, and
- treatment-by-time interactions to capture changes in impact over the study period.

Random effects were specified at both the school and pupil levels to account for clustering and individual variation in progress over time. This modelling approach enables precise estimation of the intervention's effect at each time point, while appropriately handling the correlated structure typical of longitudinal, cluster-randomised designs.

Effect Size Calculation

Treatment effects were expressed using Cohen's *d*, a standardised measure of effect size suitable for comparing group differences. This statistic allows results to be interpreted in terms of the magnitude of the intervention's impact, rather than raw score differences. Cohen's *d* represents the difference between group means divided by a suitable standard deviation. In our case, consistent with Westfall *et al.* (2014), we divided estimated mean differences by the total standard deviation of the multilevel model.

Subgroup Analyses

Pre-planned subgroup analyses explored whether the intervention had different effects for Pupil Premium (PP) and non-Pupil Premium pupils. These analyses used the same mixed-effects modelling framework as the primary analysis but were conducted separately for each subgroup. Separate analyses were chosen instead of using interaction terms to allow for distinct variance structures and random-effects patterns across groups, while maintaining sufficient statistical power to detect meaningful effects within each subgroup. For each subgroup, effect sizes and confidence intervals were calculated using Cohen's *d*. This approach provides a clear and reliable estimate of how the intervention's impact may differ between disadvantaged and non-disadvantaged pupils.

Secondary Analyses

School-level Analysis

Secondary analyses examined gain scores, calculated as the difference between each follow-up measurement and the baseline score. These analyses provided complementary insights into the intervention's effects and helped to clarify patterns of pupil progress over time. School-level summaries of sample sizes and mean gains were also produced to explore variation between schools and to identify potential differences in implementation or engagement with the intervention.

Low pre-test scores

An additional secondary analysis was conducted focusing on pupils with lower baseline performance. Low pre-test scores were defined as those below the median, calculated separately for Pupil Premium (PP) and non-Pupil Premium pupils. This approach

allowed the study to explore whether Eedi had a differential impact among pupils who started with lower levels of attainment.

CACE Analysis

To account for differences in adherence, a Complier Average Causal Effect (CACE) analysis was conducted using two-stage least squares regression within a multilevel framework to reflect the clustered trial design. Treatment randomisation served as an instrumental variable for adherence behaviour and met the standard criteria for instrument strength, independence, and monotonicity.

In the first stage, pupil adherence was modelled as a function of treatment allocation, controlling for baseline achievement (decomposed into within-school deviations and between-school means). The second stage then estimated the impact of predicted adherence on outcomes using multilevel regression with school-level random effects. Adherence was operationalised using four binary thresholds based on cumulative doses at each time point of outcome assessment: half-dose (≥ 30 , ≥ 60 , ≥ 90 , ≥ 120 units), three-quarter-dose (≥ 45 , ≥ 90 , ≥ 135 , ≥ 180 units), recommended dose (≥ 60 , ≥ 120 , ≥ 180 , ≥ 240 units), and power dose (≥ 100 , ≥ 200 , ≥ 300 , ≥ 400 units).

CACE parameters represent the causal effects among adherent subgroups, reducing potential dilution of impact that can occur in traditional intent-to-treat analyses. Effect sizes were reported using Cohen's *d*, consistent with the primary analytical framework.

Implementation and Process Evaluation

Teacher Surveys

Teachers were invited to complete an online survey to gather their experiences of using Eedi. A scale adapted from Hettinger et al., (2022) was used to explore teacher confidence in teaching mathematics. Teachers were also asked to respond to statements about pupil engagement, Eedi feature use and perceived impact on workload.

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, two 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, Feters, & 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.

Impact Results: Summary

The impact evaluation assessed the effect of *Eedi* on Key Stage 3 mathematics attainment over a 24-month period, using NWEA MAP Growth RIT scores as the primary outcome. The analysis combined intention-to-treat (ITT), subgroup, sensitivity, and compliance (CACE) models, alongside platform usage data, to explore both the magnitude of impact and the mechanisms driving outcomes.

Overall Effects

Across all time points, the estimated effects of *Eedi* were positive. The effect of intervention was significant at 18 and 24 months. The direction of effect remained consistently positive throughout the trial, increasing from small at six months to moderate by 18 months. This trend indicates that *Eedi's* potential benefit strengthened over time, suggesting that sustained use may be required for measurable learning gains to emerge.

Subgroup and Sensitivity Analyses

Effects were positive for both Pupil Premium (PP) and non-Pupil Premium pupils at all time points, with slightly larger estimates among non-PP pupils. For PP pupils, the effect peaked at 18 months, while non-PP pupils showed the largest gains at 18-24 months. A sensitivity analysis focusing on pupils with lower baseline scores showed a similar pattern: small effects at 6 and 12 months, increasing to moderate at 18 months before stabilising at 24 months. These findings suggest that *Eedi* may be beneficial even for pupils starting from a lower level of attainment. CACE analysis revealed a clear dose-response relationship between engagement and learning outcomes. By 18 months, pupils with high adherence achieved statistically significant improvements, with effect sizes approaching or exceeding $d = 1.0$ for the most compliant groups. This confirms that *Eedi's* effectiveness is strongly implementation-dependent: pupils who used the platform consistently realised larger and more sustained gains.

Impact Results: Technical Analysis

Sample Characteristics

A total of 20 schools participated in the study, contributing 3,448 students with available pre-test scores. At six months, 16 schools participated, comprising 2,326 students. Four schools were not able to participate at six months for logistical reasons. By 12 months, participation increased to 20 schools with 2,901 students. At 18 and 24 months, however, 11 schools withdrew, reducing the sample to 1,153 and 1,192 students, respectively. Additional characteristics, including mean and standard deviation of test scores as well as the number of pupils eligible for Pupil Premium, are summarised in Table 1.

Time Period	Treatment	School (N)	All pupils (N)	Mean (SD)	PP (N)	Non-PP (N)	Pupil-level (%)
<i>Pre-score (Baseline)</i>							
	Control	8	1226	211.4 (14.29)	435	791	35.48%
	Treatment	12	2222	211.18 (15.54)	623	1598	28.04%
<i>6 months</i>							
	Control	7	910	213.97 (15.6)	296	614	32.53%
	Treatment	9	1416	211.57 (14.96)	438	978	30.93%
<i>12 months</i>							
	Control	8	1032	215.79 (15.78)	350	682	33.91%
	Treatment	12	1869	216.9 (16.78)	519	1350	27.77%
<i>18 months</i>							
	Control	3	384	214.76 (16.43)	127	257	33.07%
	Treatment	6	769	216.86 (16.11)	276	493	35.89%
<i>24 months</i>							
	Control	3	437	215.38 (16.29)	161	276	36.84%
	Treatment	6	755	217.34 (17.07)	245	510	32.45%

Table 1: Participant characteristics and test scores

Attrition

Attrition in this trial has two distinct layers with different implications: school-level withdrawal between the first and second year, which accounts for most of the missing data, and individual non-response within participating schools, which was modest throughout.

School-level withdrawal was the dominant source of sample loss. Eleven of the 20 schools did not continue into the second year, for reasons recorded as logistical rather than outcome-related, and this happened largely as a block before the 18- and 24-month outcomes existed. Nine schools continued (six intervention, three control); withdrawal was somewhat higher among control schools (five of eight, 62%) than intervention schools (six of twelve, 50%).

Individual non-response within the nine continuing schools was modest: 18-month response was 79.9% in the intervention arm and 72.2% in control, and at 24 months 78.4% and 82.1%. The differential is small and changes direction between waves, which is not the signature of systematic outcome-related selection. At the trial level, response rates were also close across arms (18 months: 34.6% against 31.3%; 24 months: 34.0% against 35.6%).

Attrition was nonetheless selective on baseline attainment: pupils who left the study had higher baseline scores than those who remained (Vanacore, 2025), so the later samples skew towards lower-attaining pupils. Composition by disadvantage moved in two directions. Within-school dropout in the first year was somewhat higher among Pupil Premium pupils (18.9% compared with 11.8%), but school-level continuation raised the Pupil Premium share of the remaining sample at 18 and 24 months (Table 1). The 18- and 24-month samples are therefore lower-attaining and slightly more disadvantaged than the baseline sample.

Longitudinal Analysis

Longitudinal analysis was performed using scores at 6, 12, 18, and 24 months as outcome variables, adjusting for baseline score (between- and within-school) and treatment, with school ID and user ID as random effects.

At 6 months, the intervention effect was very small ($d = 0.06$, 95% CI: -0.1 to 0.23). At 12 months, the effect size remained small ($d = 0.17$, 95% CI: -0.003 to 0.34). By 18 months, the estimated effect size increased to a moderate level ($d = 0.46$, 95% CI: 0.24 to 0.68), suggesting a potential benefit of the intervention. At 24 months, the effect size slightly decreased but remained positive ($d = 0.3$, 95% CI: 0.11 to 0.5).

The point estimates suggest that the intervention effect increased over time, with the largest effect observed at 18 months before a partial decline at 24 months.

Time Point	Effect Size (95% CI)	P-value
6 Months	0.06 (-0.1, 0.23)	0.4239
12 Months	0.17 (-0.003, 0.34)	0.0531
18 Months	0.46 (0.24, 0.68)	0.0002
24 Months	0.3 (0.11, 0.5)	0.0037

Table 2: Longitudinal analysis

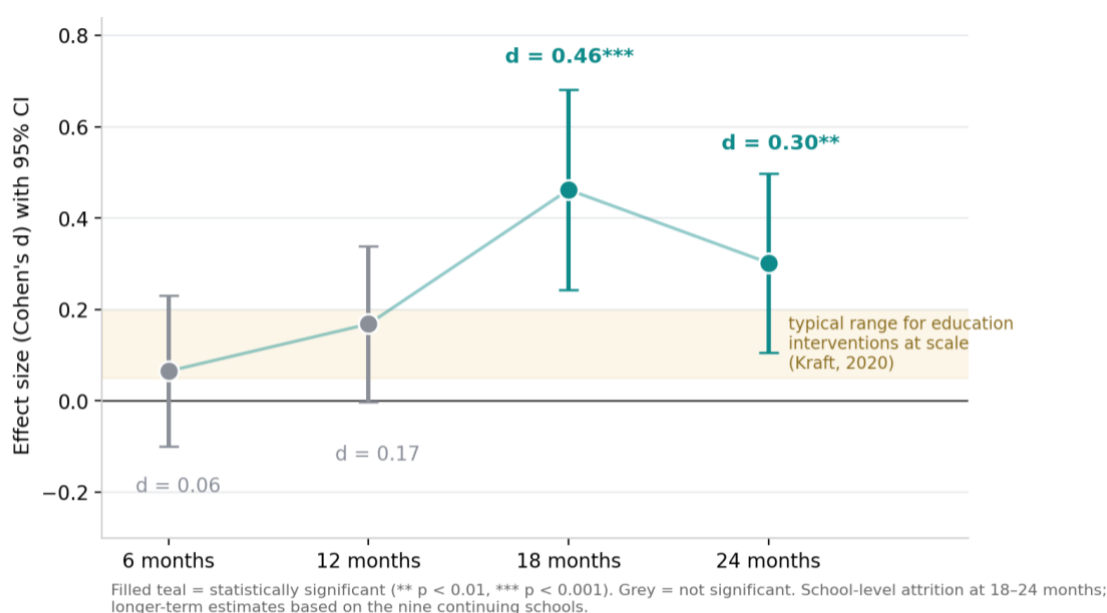


Figure 2. Intention-to-treat effects on NWEA MAP Growth at 6, 12, 18 and 24 months (Cohen's d with 95% CI). Filled markers are statistically significant (18 months $p = 0.0002$; 24 months $p = 0.0037$).

Subgroup Analysis

Subgroup analysis was conducted for PP and non-PP students (Table 3). Among PP students, effect sizes were small but positive at 6 months ($d = 0.111$, 95% CI: -0.101 to 0.323), and 12 months ($d = 0.142$, 95% CI: -0.065 to 0.348). At 18 months, the estimated effect increased to a small-to-moderate level ($d = 0.37$, 95% CI: 0.106 to 0.635), before declining again at 24 months ($d = 0.097$, 95% CI: -0.144 to 0.338).

For non-PP students, effect sizes were generally larger. Although very small at 6 months ($d = 0.034$, 95% CI: -0.151 to 0.218), the treatment group showed small positive effects at 12 months ($d = 0.171$, 95% CI: -0.018 to 0.36) and moderate effects at 18 months ($d = 0.508$, 95% CI: 0.265 to 0.751) and 24 months ($d = 0.4$, 95% CI: 0.171 to 0.629).

Time Point	Group	Effect Size (95% CI)	P-value
<i>6 Months</i>			
	Non-PP	0.034 (-0.151, 0.218)	0.7088
	PP	0.111 (-0.101, 0.323)	0.2896
<i>12 Months</i>			
	Non-PP	0.171 (-0.018, 0.36)	0.0742
	PP	0.142 (-0.065, 0.348)	0.1675
<i>18 Months</i>			
	Non-PP	0.508 (0.265, 0.751)	0.0002
	PP	0.37 (0.106, 0.635)	0.0075
<i>24 Months</i>			
	Non-PP	0.4 (0.171, 0.629)	0.0012
	PP	0.097 (-0.144, 0.338)	0.418

Table 3: Subgroup Analysis

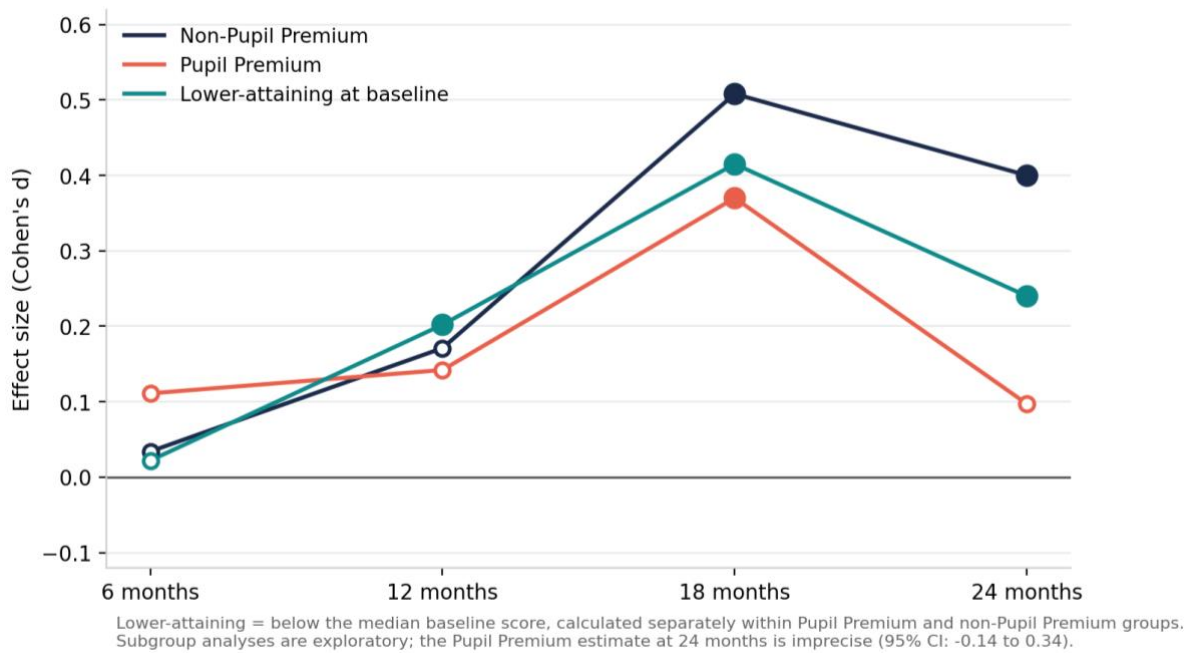


Figure 3. Intention-to-treat effects by subgroup over time (Cohen's d); filled markers are statistically significant. Lower-attaining pupils are those below the median baseline score, calculated within Pupil Premium and non-Pupil Premium groups.

Sensitivity Analysis

Low Pre-Test Scores

We conducted a sensitivity analysis including students who scored below the median for their PP group at baseline (results in table 4). After adjusting for baseline performance, students in the intervention group consistently outperformed those in the control group at 6, 12, 18, and 24 months. The largest effect was observed at 18 months, followed by 24 months, suggesting that the impact of the intervention increases over time before slightly declining. At 6 months, the difference between groups was minimal. Overall, these findings indicate that the intervention is also beneficial for students with lower baseline scores, with the most pronounced gains emerging around 18 months.

Time Point	Effect Size (95% CI)	P-value
6 Months	0.022 (-0.159, 0.203)	0.8052
12 Months	0.202 (0.012, 0.392)	0.0383
18 Months	0.415 (0.184, 0.646)	0.0009
24 Months	0.24 (0.03, 0.451)	0.0263

Table 4. Low Pre-Test Score Sensitivity Analysis

School-level Analysis

School-level analysis was performed by calculating mean gain defined as the difference between pre-score and post-score at different time points for all schools and results have been shown in table 5. Overall, both groups showed positive gains from baseline at 6, 12, 18, and 24 months, with some variation across schools.

In the control group, most schools showed gradual improvements over time, with some schools having missing data at later follow-up points. For instance, gains at 6 months were generally modest, and while some schools continued to show increases up to 12 months, data beyond this period were limited for several schools.

In the treatment group, gains were generally higher and more consistent across schools, particularly at 12 and 24 months. Some schools exhibited substantial improvements by 18 months, indicating the intervention had a positive effect over time. However, a few schools had missing data at certain follow-up points, which may affect comparisons.

Treatment	School Id	6 Months		12 Months		18 Months		24 Months	
		Sample Size	Mean gain (SD)	Sample Size	Mean gain (SD)	Sample Size	Mean gain (SD)	Sample Size	Mean gain (SD)
Control	6	111	0.57 (9.78)	122	2.28 (8.74)	NA	NA	NA	NA
	7	117	4.8 (8.35)	103	6.69 (8.26)	94	9.8 (10.01)	102	10.31 (8.05)
	11	NA	NA	75	6.51 (8.13)	NA	NA	NA	NA
	15	103	3.09 (7.51)	89	3.55 (9.83)	NA	NA	NA	NA
	16	148	2.8 (10.22)	186	5.82 (8.47)	NA	NA	NA	NA
	24	203	3.51 (10.13)	210	3.4 (11.33)	183	3.45 (14.39)	206	6.36 (12.85)
	30	122	1.39 (13.51)	132	0.9 (11.65)	107	6.04 (11.48)	129	4.94 (13.77)
	31	106	2.01 (9.62)	115	2.14 (8.94)	NA	NA	NA	NA
Treatment	1	311	5.98 (8.48)	298	7.56 (7.37)	NA	NA	NA	NA
	2	121	1.89 (9.37)	128	3.87 (10.39)	105	6.82 (10.85)	115	8.6 (12.62)
	4	155	1.31 (9.94)	168	2.58 (9.54)	165	7.38 (12.27)	148	3.92 (11.16)
	10	NA	NA	130	6.92 (9.31)	136	9.03 (10.88)	130	11.95 (11.43)
	12	169	1.74 (8.69)	161	4.71 (8.66)	152	9.63 (8.25)	150	10.65 (8.57)
	13	183	3.06 (7.47)	171	5.11 (7.91)	NA	NA	NA	NA
	17	NA	NA	137	6.55 (6.38)	NA	NA	NA	NA
	19	NA	NA	146	6.1 (8.09)	NA	NA	NA	NA
	20	90	1.79 (9.64)	94	5.32 (8.35)	94	12.47 (9.28)	92	13.22 (11.93)
	28	135	1.46 (8.36)	138	5.26 (9.27)	117	8.7 (8.63)	120	5.91 (9.4)
	29	102	3.95 (10.21)	130	5.84 (10.03)	NA	NA	NA	NA
	32	150	4.85 (8.34)	168	6.4 (9.6)	NA	NA	NA	NA

Table 5. School-level Sensitivity Analysis

CACE Analysis

CACE (Complier Average Causal Effect) analyses were conducted to examine the impact of adherence on post-intervention outcomes, with findings presented in Tables 6 to 8. At the earlier follow-up, stronger intervention effects were observed among participants meeting higher compliance thresholds, suggesting that greater engagement was associated with improved outcomes. This pattern became more pronounced at later follow-ups, where higher levels of compliance consistently corresponded to larger effect sizes. Overall, the results indicate a clear and sustained dose-response relationship, with increased adherence linked to stronger intervention effects across the full follow-up period.

Time Point	Cut-off	Number of Compliers	Percentage of Compliers	Effect Size (95% CI)
<i>6 Months</i>				
	>=30	1071	75.64%	0.08 (0 to 0.16)
	>=45	929	65.61%	0.11 (0.02 to 0.19)
	>=60	742	52.40%	0.17 (0.08 to 0.26)
	>=100	276	19.49%	0.56 (0.44 to 0.69)
<i>12 Months</i>				
	>=60	1388	74.26%	0.25 (0.18 to 0.32)
	>=90	1184	63.35%	0.31 (0.24 to 0.39)
	>=120	963	51.52%	0.42 (0.34 to 0.49)
	>=200	445	23.81%	1.04 (0.94 to 1.15)
<i>18 Months</i>				
	>=90	568	73.86%	0.58 (0.46 to 0.7)
	>=135	516	67.10%	0.63 (0.51 to 0.75)
	>=180	436	56.70%	0.74 (0.62 to 0.86)
	>=300	226	29.39%	1.45 (1.29 to 1.6)
<i>24 Months</i>				
	>=120	592	78.41%	0.37 (0.25 to 0.48)
	>=180	531	70.33%	0.41 (0.3 to 0.53)
	>=240	454	60.13%	0.48 (0.36 to 0.6)
	>=400	291	38.54%	0.84 (0.7 to 0.98)

Table 6. CACE Analysis

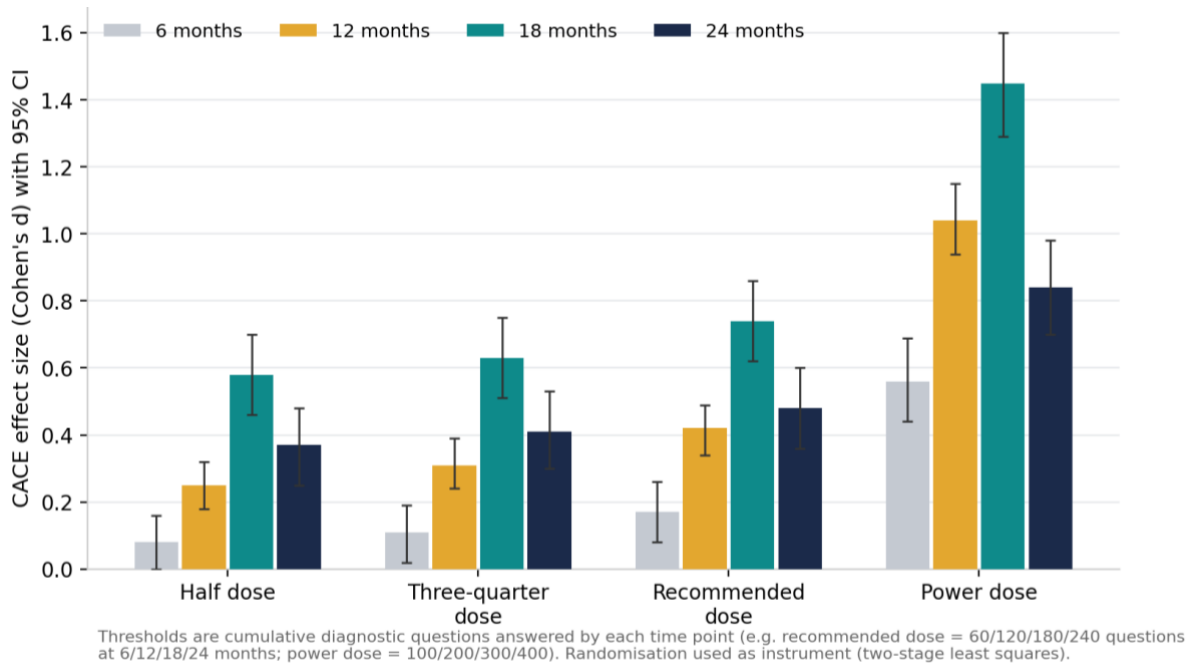


Figure 4. Complier average causal effects by engagement threshold and time point (Cohen's d with 95% CI). Thresholds are cumulative diagnostic questions answered; randomisation is used as the instrument.

Time Point	Cut-off	Number of Compliers	Percentage of Compliers	Effect Size (95% CI)
<i>6 Months</i>				
	>=30	298	68.04%	0.16 (0.01 to 0.31)
	>=45	248	56.62%	0.2 (0.05 to 0.35)
	>=60	188	42.92%	0.27 (0.1 to 0.44)
	>=100	72	16.44%	0.54 (0.29 to 0.78)
<i>12 Months</i>				
	>=60	334	64.35%	0.29 (0.15 to 0.43)
	>=90	276	53.18%	0.37 (0.22 to 0.51)
	>=120	209	40.27%	0.49 (0.33 to 0.65)
	>=200	91	17.53%	1.11 (0.89 to 1.33)
<i>18 Months</i>				
	>=90	174	63.04%	0.53 (0.33 to 0.73)
	>=135	151	54.71%	0.6 (0.39 to 0.8)
	>=180	115	41.67%	0.76 (0.54 to 0.99)
	>=300	47	17.03%	1.47 (1.15 to 1.79)
<i>24 Months</i>				
	>=120	171	69.80%	0.15 (-0.05 to 0.35)
	>=180	148	60.41%	0.18 (-0.02 to 0.38)
	>=240	116	47.35%	0.23 (0.02 to 0.45)
	>=400	62	25.31%	0.37 (0.1 to 0.64)

Table 7. CACE Analysis (PP)

Time Point	Cut-off	Number of Compliers	Percentage of Compliers	Effect Size (95% CI)
<i>6 Months</i>				
	>=30	773	79.04%	0.04 (-0.06 to 0.14)
	>=45	681	69.63%	0.07 (-0.03 to 0.17)
	>=60	554	56.65%	0.12 (0.02 to 0.23)
	>=100	204	20.86%	0.53 (0.38 to 0.68)
<i>12 Months</i>				
	>=60	1054	78.07%	0.23 (0.15 to 0.32)
	>=90	908	67.26%	0.29 (0.2 to 0.38)
	>=120	754	55.85%	0.39 (0.3 to 0.48)
	>=200	354	26.22%	1 (0.88 to 1.12)
<i>18 Months</i>				
	>=90	394	79.92%	0.6 (0.46 to 0.75)
	>=135	365	74.04%	0.64 (0.5 to 0.79)
	>=180	321	65.11%	0.73 (0.58 to 0.88)
	>=300	179	36.31%	1.39 (1.21 to 1.58)
<i>24 Months</i>				
	>=120	421	82.55%	0.47 (0.33 to 0.61)
	>=180	383	75.10%	0.51 (0.37 to 0.65)
	>=240	338	66.27%	0.58 (0.44 to 0.73)
	>=400	229	44.90%	0.97 (0.81 to 1.13)

Table 8. CACE Analysis (Non-PP)

IPE Results: Summary

The Implementation and Process Evaluation (IPE) explored how *Eedi* was implemented across participating schools, how teachers and pupils engaged with the platform, and which contextual factors shaped variation in impact. Data were collected through teacher surveys, interviews, and platform usage records, and analysed using a deductive-inductive framework structured around usability, access, engagement, content alignment, workflow, and implementation.

Implementation and Engagement

Teachers described *Eedi* as intuitive, time-saving, and easy to integrate into classroom practice. Many reported reduced marking and planning workload, improved insight into misconceptions, and more efficient tracking of pupil progress. However, the level of engagement varied across schools, influenced by access, staffing, and departmental culture.

Set-Based Engagement and Progression

Teachers consistently described an attainment-linked pattern in quiz progression. Higher-attaining sets progressed smoothly, while middle and lower sets often stalled in diagnose–resolve loops or encountered sharp increases in difficulty. This was reported as demotivating for some pupils. Teachers also highlighted the need for better curriculum mapping and for more lower-attainer content.

Connectivity, Accessibility, and Culture

Access to reliable devices and Wi-Fi was a persistent barrier. Forty-three percent of teachers cited in-school connectivity issues, and 40% reported home access challenges. Interviews described reduced device availability (e.g., from three laptop trolleys to one), IT-room booking pressure, and reliance on pupil mobile phones. Concerns around opportunities for pupils to access *Eedi* at home were also significant.

Incentives and Motivational Features

Teachers identified the £30 monthly voucher as a strong motivator, particularly when promoted across the department. Other motivational features such as avatars, games, and leaderboards were appreciated but inconsistently used due to firewall restrictions, parental sign-up requirements, or cultural fit. Teachers also described using progress displays, class competitions, and quiz targets to sustain engagement.

What teachers told us

Implementation and process evaluation: two-year cluster randomised controlled trial of Eedi

WHAT WORKED

Less marking, more insight

Teachers described Eedi as intuitive and time-saving, with reduced marking and planning workload. 80% used the data to review misconceptions, 77% to inform planning and 93% to set quizzes.

Easy to embed in classroom practice

Teachers found Eedi straightforward to integrate into weekly routines, with more efficient tracking of pupil progress and clearer insight into individual misconceptions (77%).

Incentives that landed

The £30 monthly voucher was a strong motivator when promoted across the department; progress displays, class competitions and quiz targets sustained engagement.

WHAT GOT IN THE WAY

Devices and connectivity

43% of teachers cited in-school connectivity issues and 40% home access challenges; interviews described laptop trolleys cut from three to one and reliance on pupil phones.

Progression pitch and feedback loops

Higher-attaining sets progressed smoothly; middle and lower sets often stalled in diagnose-resolve loops or met sharp difficulty ramps, which some pupils found demotivating.

Homework culture and staffing

Weak homework cultures (60 to 65% completion even in top sets) and understaffed departments meant some schools slipped from weekly to fortnightly use.

HOW TEACHERS USED THE DATA

Share of surveyed teachers using Eedi data for each purpose



ACCESS BARRIERS TEACHERS REPORTED



TEACHER VOICES

"My pupils loved using Eedi and were highly motivated to use it every week."

Survey free-text response

"I always set it diagnose and... they get a question wrong and then they end up in that kind of cycle... they constantly get stuck."

Class teacher, on middle sets

"We used to have three laptop trolleys... we've gone down to one... getting access is hard."

Class teacher, on device constraints

"The £30 Amazon voucher... massively increased [engagement]... more than six quizzes per lesson."

Class teacher, on incentives

Make it weekly. Make it retrieval.

Sustained, consistent use is where the gains were: average effects reached $d = 0.46$ at 18 months and $d = 0.30$ at 24 months, and pupils meeting the highest engagement threshold reached $d = 1.04$ at 12 months and $d = 1.45$ at 18 months.

Source: WhatWorked Education, two-year school-randomised cluster RCT of Eedi (20 schools, 3,448 pupils at baseline; NWEA MAP Growth outcome). IPE: 49 completed teacher surveys (30 intervention, 19 control) and 2 follow-up interviews.

Figure 5. Implementation and process evaluation: what worked, what got in the way, and what teachers reported in this trial.

IPE Survey Results: Technical Analysis

Teacher Characteristics

Across groups, the sample was predominantly mathematics specialists with a broad range of teaching experience. Of 11 control group teachers: 2 were senior leaders, 2 head of department, 2 assistant head of department, 7 class teachers. 9 control group teachers identified as specialist maths teachers, 2 as PE teachers. Of 30 intervention group teachers: 1 was a senior leader, 2 head of department, 5 assistant head of department, 22 class teachers. 28 identified as specialist maths teachers, with 1 PE teacher and 1 primary teacher.

Teaching experience was predominantly extensive in both groups. In the control group, nearly half had 10+ years' experience (46%), with a further 36% in the 6-10 band; only 9% each were in the 1-2 and 3-5 bands. The intervention group reflected a similar profile, with 43% in the 10+ category and 30% in 6-10, alongside a larger mid-career contingent (3-5 years: 20%) and a small early-career group (1-2 years: 7%). Overall, 82% of control and 73% of intervention teachers had at least six years' experience, indicating broadly comparable, and generally experienced, samples.

Teaching Condition	Years Teaching Experience	N	%
<i>Control</i>	1-2	1	9%
	3-5	1	9%
	6-10	4	36%
	10+	5	46%
<i>Intervention</i>	1-2	2	7%
	3-5	6	20%
	6-10	9	30%
	10+	13	43%

Table 9. Years teaching experience across IPE survey respondents

Teacher Confidence

A scale adapted from Hettinger et al., (2022) was used to explore teacher confidence in teaching mathematics. On all confidence items, responses reflected certainty/confidence, with the strongest agreement for identifying support needs and differentiating in maths, and more moderate confidence when motivating students who often fail school requirements.

Helping pupils value learning in maths

Most teachers reported high confidence in helping pupils value learning in mathematics. In the control group, no respondents selected options below “Slightly Certain,” and nearly three-quarters were “Very Certain” (73%). In the intervention group, confidence was also high, with almost two-thirds “Very Certain” (63%); a small minority reported being “Very Uncertain” (3%) or “Neutral” (7%). Overall, responses in both arms are skewed toward certainty.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control (N)</i>	0	0	0	3	8
<i>Control (%)</i>	0%	0%	0%	27%	73%
<i>Intervention (N)</i>	1	0	2	8	19
<i>Intervention (%)</i>	3%	0%	7%	27%	63%

Table 10. Responses to ‘How certain are you that you can help pupils value learning in maths?’

Motivating students with low interest in maths

Most teachers reported confidence in motivating students with low interest in maths. In the control group, around two-thirds felt very certain (64%) and a further 18% slightly certain, with only small proportions selecting slightly uncertain (9%) or neutral (9%). In the intervention group, no teachers selected either uncertainty option; responses were concentrated in slightly certain (47%) and very certain (43%), with 10% neutral. Overall, the distribution in both arms is skewed towards certainty.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control</i>	0	1	1	2	7
<i>Control (%)</i>	0%	9%	9%	18%	64%
<i>Intervention</i>	0	0	3	14	13
<i>Intervention (%)</i>	0%	0%	10%	47%	43%

Table 11. Responses to ‘How certain are you that you can motivate students who show low interest in maths?’

Motivating students who often fail school requirements

Confidence in motivating students who often fail school requirements was more moderate. In the control group, most respondents were slightly certain (46%) or very certain (36%), though a small minority reported being very uncertain (9%) or neutral (9%). In the intervention group, responses clustered around slightly certain (44%), with

notable proportions neutral (23) and very certain (20%); a smaller share indicated slight (10%) or significant uncertainty (3%).

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control</i>	1	0	1	5	4
<i>Control (%)</i>	9%	0%	9%	46%	36%
<i>Intervention</i>	1	3	7	13	6
<i>Intervention (%)</i>	3%	10%	23%	44%	20%

Table 12. Responses to 'How certain are you that you can motivate even those students who often fail school requirements?'

Identifying students who need support across maths topics

Confidence in diagnosing support needs across maths topics was high. In the control group, nearly three-quarters reported being very certain (73%), with a further 18% slightly certain; only 9% indicated slight uncertainty and none selected neutral or very uncertain. In the intervention group, 60% were very certain and 33% slightly certain, with a small minority neutral (7%) and no respondents reporting uncertainty. Overall, responses in both arms are strongly skewed toward certainty.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control</i>	0	1	0	2	8
<i>Control (%)</i>	0%	9%	0%	18%	73%
<i>Intervention</i>	0	0	2	10	18
<i>Intervention (%)</i>	0%	0%	7%	33%	60

Table 13. Responses to 'How certain are you that you can identify which students need support across different topics in maths?'

Differentiating to meet diverse needs in maths

Confidence in differentiating for diverse needs in maths was high across both arms. In the control group (n=11), nearly two-thirds were very certain (64%) and a further 27% slightly certain, with only 9% slightly uncertain. In the intervention group (n=17), responses were also skewed toward certainty, with slightly certain (53%) and very certain (41%) accounting for the vast majority; only 6% selected slightly uncertain, and no respondents chose very uncertain or neutral.

Condition	Very Uncertain	Slightly Uncertain	Neutral	Slightly Certain	Very Certain
<i>Control</i>	0	1	0	3	7
<i>Control (%)</i>	0%	9%	0%	27%	64%
<i>Intervention</i>	0	1	0	9	7
<i>Intervention (%)</i>	0%	6%	0%	53%	41%

Table 14. Responses to 'How certain are you that you can differentiate to meet the diverse needs of your students in maths?'

Using technology to enhance maths learning

Confidence in using technology to enhance maths learning was high in both arms. In the control group, over half reported being very confident (55%) and a further 36% slightly confident, with only 9% neutral and no reports of low confidence. In the intervention group, responses were similarly concentrated at the confident end of the scale: slightly confident (47%) and very confident (40%) together accounted for 87%, with small proportions neutral (10%) or slightly unconfident (3%), and none very unconfident.

Condition	Very Unconfident	Slightly Unconfident	Neutral	Slightly Confident	Very Confident
<i>Control</i>	0	0	1	4	6
<i>Control (%)</i>	0%	0%	9%	36%	55%
<i>Intervention</i>	0	1	3	14	12
<i>Intervention (%)</i>	0%	3%	10%	47%	40%

Table 15. Responses to 'How confident are you that using technology can enhance student learning in maths?'

Maths Homework

Time Expectations and Completion Rates

Most teachers in both groups reported setting maths homework. Average monthly time expectations were similar (78 minutes control; 80 minutes intervention), with the intervention group reporting a wider range in both monthly minutes and number of pieces set. Of the 11 control group teachers, 8 stated that they set maths homework with their class. Of the 30 intervention group teachers, 27 stated that they set maths homework with their class.

For control group teachers, time per month students were expected to spend on completing maths homework ranged from 20-120 minutes, averaging 78 minutes per month. Students completing 1-4 pieces of homework a month, average expected task time 29 mins (expected time per task 20-40 minutes).

For intervention teachers, time per month students were expected to spend on completing maths homework ranged from <10-180 minutes, averaging 80 minutes per month. Students completing 1-16 pieces of homework a month, average expected task time 19 minutes (expected time per task <10-45 minutes).

Completion was generally moderate to high in both arms. In the intervention group, nearly three-quarters (74%) reported completion at 50% or higher, with the largest share in the 50–75% band (41%). The control group showed a similar pattern, with 75% reporting 50% or higher completion and half of respondents in the 50–75% band (50%). Very low completion (<25%) was uncommon in both groups (intervention: 11.1%; control: 12.5%).

Condition	Completion Rate	N	%
<i>Intervention</i>	Less than 25%	3	11%
	25-50%	4	15%
	50-75%	11	41%
	75%+	9	33%
<i>Control</i>	Less than 25%	1	12.5%
	25-50%	1	12.5%
	50-75%	4	50%
	75%+	2	25%

Table 16. Homework completion rates in intervention and control conditions

Platform Use

Eight control-group teachers reported regularly using an online platform to assign maths homework (Dr Frost n=2, Sparx n=1, Eedi n=1, MyMaths n=6, Desmos & Transum n=1, Quizizz n=1, Khan Academy n=1). Twenty-five intervention-group teachers reported regular platform use; 18 were using the intervention platform, Eedi. Of those using Eedi, 12 also used additional platforms (Dr Frost n=7, Mathswatch n=1, MyMaths n=3). Seven intervention teachers were not using Eedi to assign homework and instead used alternative platforms (Mathswatch n=6, MyMaths n=1).

Monitoring homework via Eedi - Teacher experience

Intervention teachers who used Eedi to set homework rated how straightforward it was to monitor homework completion and performance. Overall, responses skewed positive: 58% found monitoring straightforward (37% very; 21% slightly), 16% were neutral, and 26% reported some level of challenge (5% very; 21% slightly).

Rating	N	%
Very Challenging	1	5%
Slightly Challenging	4	21%
Neutral	3	16%
Slightly Straightforward	4	21%
Very Straightforward	7	37%

Table 17. Teacher experiences of monitoring homework using Eedi

Monitoring homework - control group teacher experience

Control-group teachers described their experience of monitoring homework completion and performance as broadly manageable. Half (50%) found it straightforward (25% very; 25% slightly), while the remaining 50% reported either neutral (25%) or slightly challenging (25%) experiences. Notably, no teachers rated the process as very challenging.

Rating	N	%
Very Challenging	0	0%
Slightly Challenging	2	25%
Neutral	2	25%
Slightly Straightforward	2	25%
Very Straightforward	2	25%

Table 18. Control group teacher experiences of monitoring homework

Eedi Classwork

Most intervention teachers reported using Eedi during lessons (n=21, 70%), indicating routine in-class integration alongside other teaching practices. A smaller group did not use Eedi in class (n=9, 30%), suggesting either differing pedagogical approaches or contextual constraints.

Monitoring classwork completion and performance

Among teachers using Eedi in class, perceptions of monitoring were generally positive. Around one-third rated it very straightforward, with a further group finding it slightly straightforward. A sizeable minority were neutral or found it slightly challenging.

Rating	N	%
Very Challenging	0	0%
Slightly Challenging	5	24%
Neutral	6	29%
Slightly Straightforward	3	14%
Very Straightforward	7	33%

Table 19. Teacher experiences using Eedi to monitor classwork

Eedi Functions

To understand how the platform was integrated in lessons, 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: 9 teachers (43%) used it to surface whole-class misconceptions, and 12 (57%) did not. Among the 12 non-users, 8 (67%) were unaware of the feature, and 4 (33%) knew about it but chose not to use it.

Diagnostic Questions

Just over half of intervention teachers reported using Eedi's diagnostic questions in classroom teaching (n=12, 57%), while 9 (43%) did not.

Eedi Resources

Use of Eedi's ancillary resources was relatively low: 4 intervention teachers (19%) reported using resources (e.g., videos, worksheets) in classroom teaching, while 17 (81%) did not.

Live Human Tutor

18 intervention teachers submitted responses to provide feedback on the live human tutor feature. Most users found the tutor somewhat helpful (n=13, 72%), with a smaller group rating it very helpful (n=4, 22%). Only one response was neutral, indicating broadly positive perceived value among those who tried it. The teacher who provided

a neutral response commented 'I know some of my students use the feature, but it's very hard for me to evaluate the impact on their learning'. Most teachers felt they had not yet used the feature enough to judge impact (n=14, 78%), with 2 teachers reporting no noticeable impact on teaching practice. Where effects were noted, teachers most commonly reported greater student independence (n=11, 61%) and, to a lesser extent, reduced time answering individual questions (n=2, 11%) and more focus on higher-level concepts (n=1, 6%).

Lesson Use

To understand how Eedi was used in lessons, 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

Most teachers used Eedi to review misconceptions with the whole class (n=24, 80%), although the frequency of this practice varied (rarely n=9, monthly n=8, weekly n=5, almost every lesson n=2). 6 teachers (20%) reported that they never used Eedi in this way.

Informing planning for the next lesson

A large majority of teachers reported having consulted Eedi to shape subsequent lessons (n=23, 77%). Again, the frequency of this practice varied (rarely n=8, monthly n=7, weekly, n=7, almost every lesson n=1). 7 teachers (23%) reported that they never used Eedi in this way.

Creating small groups for targeted support

Half of the intervention teachers (n=15, 50%) reported that they had used Eedi to create small groups of students for targeted support during lessons (rarely n=8, monthly n=4, weekly n=2, almost every lesson n=1). 15 teachers reported that they had never used Eedi in this way.

Identifying individual student misconceptions

Most teachers used Eedi to support individual follow-up (n=23, 77%), with usage spread across rarely (n=8), monthly (n=6), and weekly (n=6); a small number used it almost every lesson (n=2) or every lesson (n=1). 7 teachers (23%) reported that they had never used Eedi in this way.

Sharing student responses to facilitate class discussion

Two-thirds of teachers shared student work via Eedi to prompt discussion (n=20, 67%), most doing so occasionally rather than routinely (rarely n=10, monthly n=6, weekly n=3, almost every lesson n=1). 10 teachers reported that they had never used Eedi in this way.

Quizzes

Nearly all teachers reported using quizzes (n=28, 93%), with only 2 intervention teachers reporting that they did not use the quizzes in Eedi. Of those teachers that used Eedi quizzes, 6 felt there was little alignment with their curriculum (21%), 16 moderate alignment (58%), 6 complete alignment (21%). Over half felt the topic quizzes were most beneficial (n=18, 60%), 6 teachers felt the retrieval practice quizzes were most beneficial (20%), with 6 teachers stating that they were unsure about the various quiz types (20%).

The target of the Eedi programme was for teachers to set approximately one quiz per week for their students. Against this target, most teachers reported consistently setting one or more quizzes weekly (n=17, 57%), some reported aiming for one per week, but often ended up setting fewer (n=10, 33%), while a small group intentionally set fewer (n=3, 10%). The most frequently cited constraints to quiz setting were concerns about student access to devices/internet at home (n=11, 37%) and low student engagement/completion (n=9, 30%), with smaller numbers highlighting curriculum pacing (n=4, 13%), time for planning/administration (n=4, 13%), data insights not sufficient to justify the time (n=4, 13%) or preference for other tools or paper-based homework (n=6, 20%). Very few mentioned technical issues (n=1, 3%), and none cited school/department priorities as a barrier.

Teachers were asked to share their thoughts on Eedi's quiz types. Eight open-ended responses were submitted, with teachers reporting strong pupil engagement, but also highlighting needs for better differentiation, tighter curriculum alignment, improved question quality and variety, and more efficient workflow/analytics (including data integrity). Several also suggested usability enhancements, such as stronger search and a more age-appropriate presentation for older year groups.

1	I need to explore EEDI more. Clearly I do not know it well enough to use it effectively.
2	More differentiation for lower ability students
3	Tracking what homework students have or haven't completed is very difficult. Have to click into each homework separately. Being able to download a spreadsheet with this information would be very helpful in monitoring homework completion. Also struggle for tasks for groups that are low attainers. Students often just click they have resolved the misconception when they have not giving false data.
4	It would be beneficial to have more quizzes aligned to the Northern Ireland curriculum.
5	I would like to see an improvement in the quality of the questions, and easier search functions.
6	Pupils engage really well with it and look forward to that period each week.
7	They were good but there could be a bigger range of questions available. Different style of questions for the topics would be beneficial and also some topics that are not available.
8	I'd love to see more quizzes with just one outcome, and geared towards the National Applications of Mathematics (Nat 4 and Nat 5). Possibly a slightly more grown up look for S3/4 age groups.

Table 20. Teacher free-text responses on Eedi quizzes

Across the eight open-ended responses, teachers described Eedi as a product that pupils enjoy using, alongside clear opportunities to improve differentiation, curriculum fit, question design/coverage, and the day-to-day workflow for monitoring learning. Although the sample is small, some themes recur across multiple responses.

Question quality, variety, and design

Teachers asked for richer item banks, alternative formats (including single-outcome quizzes), and a more mature look for older cohorts:

"I would like to see an improvement in the quality of the questions, and easier search functions."

"There could be a bigger range of questions ... different style of questions ... and also some topics that are not available."

"I'd love to see more quizzes with just one outcome ... a slightly more grown up look for S3/4."

Differentiation and support for lower attainers

There were some calls for more scaffolded content and tasks for struggling learners:

"More differentiation for lower ability students."

"Struggle for tasks for groups that are low attainers."

Curriculum alignment and coverage

There were some requests to extend/clarify mapping for Northern Ireland and Scottish Nat 4/Nat 5, and broaden topic coverage:

"More quizzes aligned to the Northern Ireland curriculum."

"More quizzes ... geared towards the National Applications of Mathematics (Nat 4 and Nat 5)."

Usability, search, and discoverability

There were also some requests for improvements to search and navigation; age-appropriate presentation:

"Easier search functions."

"A slightly more grown up look for S3/4 age groups."

Workload

Time spent on lesson planning/preparation and homework

Intervention teachers 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 2-700 minutes, with an average of 215 minutes per week. Control teachers also reported wide variability, from 4-600 minutes, averaging 159 minutes per week. Teachers were also asked to estimate on average how many minutes they spend on homework setting/marking per week. Intervention teachers estimated 2-300 minutes per week on homework tasks, with an average of 54 minutes. Control teachers reported 0-120 minutes per week, averaging 39 minutes. On average, intervention teachers reported more time on both planning (+56 minutes/week) and homework-related tasks (+15 minutes/week) than control teachers, although the broad ranges, particularly for planning, suggest substantial differences in individual or school contexts.

Perceived impact of Eedi on workload

Intervention teachers were asked how they felt using Eedi had affected their workload. A majority perceived some degree of workload reduction when using Eedi (n=16, 53%), while some experienced no change (n=12, 40%). A minority reported a slight increase (n=2, 7%).

Reported Impact	N	%
Significantly decreased	2	7%
Somewhat decreased	14	46%
Neither increased or decreased	12	40%
Somewhat increased	2	7%
Significantly increased	0	0%

Table 21. Perceived impact of Eedi on teacher workload

Teachers most frequently reported time decreases for marking/correcting pupils' work and individual planning/preparation. More modest reductions were noted for assessments, general admin, and organising resources.

Reports of time increases were rare and appeared to be related to organising resources and planning; no teachers reported increases for marking, pupil supervision, teamwork within school, or staff meetings. Overall, the balance of responses suggests a net workload benefit concentrated around marking and planning.

Task	Decreased time (N)	Increased time (N)
Individual planning or preparation of lessons	10	1
Marking/correcting pupils work	17	0
Planning, administering and reporting on pupil assessments	8	1
Pupil supervision and tuition outside of timetabled lessons	3	0
General administrative work	8	1
Organising resources	6	2
Communication and cooperation with parents or guardians	1	1
Teamwork and dialogue with colleagues within this school	1	0
Staff meetings	0	0

Table 22. Workload increase/decrease associated with Eedi use

Pupil Engagement

Most intervention group teachers reported stable or improved pupil engagement with Eedi; few reported declines. Specifically: 9 reported that engagement had increased (30%), 4 significantly increased (13%), 10 neither increased nor decreased (33%), 5 didn't know (17%), and 2 reported decreased engagement (7%). When asked whether PP/FSM pupils required more support than non-PP/FSM pupils to use Eedi, responses were mixed. Ten teachers reported the same amount (34%), 9 slightly more (30%), 1 significantly more (3%), 1 slightly less (3%) and 9 didn't know (30%).

Five free-text comments were provided by teachers when asked for further comments on student engagement. They cluster around two themes: access and setup support, and engagement with specific content formats.

1	More support in having technology to complete the homework
2	Students don't like to watch the videos or do the practice questions
3	Using has decreased by time setting up homework and the administration associated with it
4	My pupils loved using Eedi and were highly motivated to use it every week.
5	FSM pupils need more support in the setup with Eedi, but once familiar they can use like other students. They need more support with the questions, but not the user/interface.

Table 23. Teacher free-text responses on student engagement with Eedi

Access and setup support

Two teachers identified barriers related to technology access at home and initial onboarding for PP/FSM pupils:

"More support in having technology to complete the homework"

"FSM pupils need more support in the setup with Eedi, but once familiar they can use like other students. They need more support with the questions, but not the user/interface."

Engagement with content formats

Reports of student engagement with content diverged: one teacher reported high motivation and weekly anticipation, while another noted resistance to videos/practice questions.

"My pupils loved using Eedi and were highly motivated to use it every week."

"Students don't like to watch the videos or do the practice questions."

Student IT issues while using Eedi

Teachers were asked: *"What are the most common IT/tech issues faced by students while using Eedi?"*. The most frequently cited problems were lack of devices and/or connectivity at school (n=13, 43%) and at home (n=12, 40%). A smaller proportion reported no issues (n=4, 13%), and one teacher noted tasks not always registering as completed (n=1; 4%).

Implementation

Training Attendance

Most teachers reported attending an Eedi training event (n=20, 67%), with a smaller proportion unsure (n=2, 7%) or not attending (n=8, 26%).

Set-Up Experience

Overall, set-up was more often described as straightforward than challenging.

Rating	N	%
Very Challenging	1	3%
Somewhat Challenging	1	3%
Neither Challenging nor Straightforward	9	30%
Somewhat Straightforward	9	30%
Very Straightforward	10	34%

Table 24. 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. Access and infrastructure issues (home device/internet access; in-school IT reliability) and student motivation were most commonly identified as significant barriers, while planning time and interpreting analytics were less frequently cited.

Barrier	Significant Barrier N (%)	Minor Barrier (N, %)	Not a Barrier (N, %)
Time required for lesson planning with Eedi	2 (7%)	3 (10%)	25 (83%)
Lack of alignment between Eedi quizzes and my curriculum sequence	2 (7%)	12 (40%)	16 (53%)
Student lack of access to devices or internet at home	7 (23%)	10 (33%)	13 (44%)
Unreliable internet or IT equipment within school	9 (30%)	7 (23%)	14 (47%)
Lack of student motivation to complete Eedi assignments	8 (27%)	6 (20%)	16 (53%)
Difficulty interpreting student results/analytics provided by Eedi	4 (13%)	7 (23%)	19 (64%)

Table 25. Reported barriers to implementation of Eedi

Teachers had the opportunity to identify further barriers through a free-text response. In total, 24 comments were submitted. Overall, they highlighted a small number of recurring pain points, led by content fit (curriculum coverage, level, and item design),

followed by access/infrastructure constraints. Workflow/analytics needs and pupil engagement factors also featured, alongside a few usability/onboarding issues.

Content alignment, coverage & design

The most commonly reported barrier was finding the right content, both in curriculum mapping (Northern Ireland; Scottish Applications of Mathematics) and in level/format for lower attainers. Teachers also noted item progression (a jump in difficulty in diagnostic items), desire for more prepared quizzes, greater topic coverage, different question styles, and the ability to assign more than five questions when needed.

"Finding tasks where every question is appropriate for my class at the stage they are in learning a given topic"

"Occasionally finding appropriate quizzes"

"The biggest issue I had was finding quizzes to match the NI curriculum... very limited quizzes aimed at lower ability levels."

"Not finding assignments to suit my needs."

"Sometimes I would like to assign more than 5 questions to certain classes."

"The style of questions available"

"Not enough lower ability numeracy recovery quizzes"

"...if there was a Scottish curriculum set (particularly Applications of Maths courses)... this would be very helpful."

"I found the 4th or 5th question on diagnostic tests jumped to very complex and this was too much for some pupils."

Access & infrastructure

Constraints in school IT (limited devices, Wi-Fi/signal issues, booking ICT rooms) and home access were frequently cited:

"Access to laptops in school"

"School IT provision"

"Cannot use in-school due to lack of IT facilities"

"No signal/Wi-Fi in my classroom... ICT rooms only have 20 computers so 10 pupils would have to work on phones... problems with Wi-Fi."

"Students having access to a suitable device... supportive parents... strong enough school culture around homework."

Workflow & analytics

Comments suggest a desire for an easier/more accessible view of homework engagement, reset for rushed attempts, and simpler analytics, including access to historical results.

"Not being able to see overall engagement of homework... having to open up each homework... not being able to re-set homework if students rush through it..."

"Analysing results"

"Analysing results, especially trying to look at previous results"

Engagement & culture

A small number of comments referenced pupil engagement directly and the need for parental support and a homework culture:

"Engaging students" / "pupil engagement" / "Engagement"

"Students having supportive parents at home... strong enough school culture around homework."

Usability & onboarding

Two comments pointed to user-friendliness and confidence with the platform:

"Didn't find that the platform was user friendly"

"I do not know the application well enough"

IPE Interview Results

In addition to completing surveys, 2 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 to surface patterns that were not pre-specified in the survey or logic model. 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 Theme	Emergence (data triggers)	Key excerpts
Diagnose-resolve loop (“getting stuck”)	Recurrent reports of pupils entering a diagnose→repair cycle after a wrong answer, especially in middle/lower sets; a distinct mechanism beyond general engagement.	<p>“I always set it diagnose and... they get a question wrong and then they end up in that kind of cycle... they constantly get stuck.”</p> <p>“The middle set gets stuck... trying to fix it...”</p>
Progression	Teachers link stalling to a quick ramp-up in difficulty and early move into problem-solving; distinct from the loop itself. Desire for fluency vs problem-solving tracks	<p>“It can ramp up in difficulty or go into a more problem-solving context quite quickly.”</p> <p>“Top set love it... will rattle through the quizzes.”</p> <p>“Some prefer... more questions on the same topic.”</p> <p>“Maybe have fluency quizzes and a problem-solving challenge quiz.”</p>
Differentiation	Persistent contrast: different sets struggle with pitch/progression; Local curriculum pathways & mapping (Scottish Applications vs Maths); leads to teachers combining platforms to secure more practice on the same topic;	<p>“Top set rattle through... middle set struggle.”</p> <p>“Teachers of the lower sets struggle... that questions are appropriately pitched.”</p> <p>“Two strands... Applications (finance-based; no algebra/trig) vs Maths (traditional STEM).”</p>

	sometimes prefers alternatives for volume	<p>"Aim to place pupils on National 5 Maths or National 5 Applications."</p> <p>"Middle ability class... White Rose... Mastery... ED... to get them a couple of rounds..."</p> <p>"I like [other platform]... can set more practice on a similar type."</p>
Device Accessibility	Accounts of reduced laptop trolleys, booking pressure, and ageing/broken devices constraining in-class use and homework, or students resorting to mobile phone use.	<p>"We used to have three laptop trolley... we've gone down to one... getting access is hard." / "One trolley with 30 laptops... another with 11... hence fortnightly rather than weekly."</p> <p>"If classroom management is right, pupils respect you re phones." / "They use their mobiles... may switch to Snapchat, but generally use it for the right reasons."</p>
Staffing	Department under-staffed; engagement framed as compliance; difficulty moving pupils between classes	<p>"We've been under staff for a long time... just trying to get it done."</p> <p>"I've struggled to move students between classes..."</p>
School Culture	Weak homework culture; long period with no consequence; even top set ~60–65% completion. Leaderboards as weak cultural fit	<p>"No consequence for homework... not a strong culture... even with my top set, only 60–65%."</p> <p>"Leaderboards... don't really happen in our school."</p>
Incentives	Scheduled retrieval slot (e.g., Thursday) to drive use, £30 Amazon voucher comms well received and associated with noticeable increases in quizzes completed; teachers actively broadcast the incentive, queries about certificates, clear targets ("complete more quizzes") and live displays used to motivate and avoid stalling.	<p>"On Thursday we really target ED and use the retrieval work... biggest buy-in."</p> <p>"The £30 Amazon voucher... massively increased [engagement]... more than six quizzes per lesson." / "We emailed out the poster... forwarded to all students."</p> <p>"Is there a certificate system we could utilise?"</p> <p>"Get them to complete more quizzes... otherwise they weren't pushing on."</p> <p>"Every week/fortnight we put up the screen... overview... it became an incentive."</p>
Technical Blocks	Potential motivators blocked e.g. avatar customisation blocked by parent sign-up, games mode blocked by firewall	<p>"They love customising their avatar, but get frustrated if the parent hasn't signed up." / "I'm desperate to get rid of that... we stop them."</p> <p>"Class like the games section... 'Complete this many quizzes then you can go on that.'... now the firewall blocks most of them."</p>

Table 26. Inductive coding of interview transcripts

Framework Analysis

After defining the inductive codes above, we created a deductive-inductive codebook (Appendix) 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 3 major themes that were explored in relation to survey and RCT results.

Set-Based Engagement & Progression

Teachers described a clear attainment-linked pattern: high-attaining sets “rattle through” quizzes, while middle/lower sets can stall when diagnostics escalate or when pupils enter a repetitive fix-it cycle when they have answered an item incorrectly. The mechanism described is twofold: (i) progression pitch (diagnostic items ramping “quite quickly” into problem-solving) and (ii) feedback cycle friction (pupils get “stuck” trying to get back on track after an error). Some pupils can find this demotivating.

“I have a top set and then a kind of middle set. And I feel like the top set love it... whereas the middle set... it can ramp up in difficulty or kind of go into a more problem-solving context quite quickly... I always set it diagnose and... they get a question wrong and then they end up in that... cycle... they constantly get stuck.”

“They’re looking for a shortcut all the time... they get a lot of questions wrong and they constantly get stuck.”

Teachers requested better curriculum mapping (e.g., Northern Ireland, Scottish Applications of Mathematics) and more content for lower attainers. Some asked for the ability to assign >5 questions and for alternative quiz formats for older cohorts.

“The biggest issue... was finding quizzes to match the NI curriculum... very limited quizzes aimed at lower ability levels... sometimes I would like to assign more than 5 questions... if there was a Scottish curriculum set... that would be very helpful... the 4th or 5th diagnostic question jumped to very complex.”

This aligns with survey free-text comments about “jumps” in diagnostic difficulty and requests for more material for lower attainers (e.g., “very limited quizzes aimed at lower ability levels”; “the 4th or 5th question on diagnostic tests jumped to very complex”).

In the main ITT analysis of the RCT, the intervention effect was positive and statistically significant overall ($d \approx 0.25$, 95% CI: 0.07 to 0.43). Subgroup exploration suggested a positive effect for both non-PP and PP students, while a sensitivity analysis with lower pre-test pupils also showed a positive effect of intervention. Teachers consistently described a pattern in which higher-attaining groups progressed smoothly through quizzes, while middle and lower sets were more likely to stall when diagnostic items escalated in difficulty or when pupils became trapped in a repetitive feedback cycle after making an error. These mechanisms, progression pitch (rapid movement into problem-solving questions) and feedback cycle friction (difficulty recovering after an error), were seen as demotivating for some pupils and may have suppressed overall impacts.

Connectivity, Accessibility and Culture

Interviews suggest that practical access to devices/Wi-Fi and room bookings materially shaped dosage. One school reported moving from three laptop trolleys to one, making fortnightly in-class use and homework setting harder. Teachers described booking pressure old/broken devices and having to resort to pupils using their own mobile devices. Weak homework cultures (limited consequences for non-completion) further constrained uptake, even in top sets.

"We used to have three laptop trolleys... we've gone down to one for the school... I used to book it weekly or certainly every other week... [this year] I probably [have] not used as much in lesson... simply because... getting access... is hard."

"They're also quite old... some coursework subjects have commandeered some... technology have trashed about three trolleys."

"Up until very recently, there was no consequence for... homework... we don't have a very strong culture of homework. So even with my top set, we're probably only looking at 60-65% completion."

These findings converge with survey responses. The most common IT problems were in-school devices/connectivity (43%) and home access (40%), and teachers explicitly tied engagement to parental support and homework culture.

Interviews suggest that departments under staffing pressure tended to treat Eedi use as another task to "get done," relying on informal internal championing (e.g.,

reminders, posters) and describing plans for alignment that could not be resourced which reduced usage to fortnightly rather than more frequent engagement.

"We... have been understaffed for a long time. It's part of it... just trying to... get it done."

Incentives and motivational features

In interviews, a monthly £30 voucher was identified as boosting engagement, especially when communicated department-wide. A number of motivational features were highlighted in interviews, but participants indicated these could be improved or were sometimes subject to limitation. These included: avatars (frustration where parent sign-up is required), games (popular but often blocked), and leaderboards (cultural fit varied).

"The £30 Amazon voucher... massively increased [engagement]... in my other year, they were completing more than six quizzes per lesson."

"They also love customising their avatar, but... get frustrated that they can't if the parent hasn't signed up... it's like... we're trying to get them to earn coins and be motivated and then we kind of stop them."

"My class like the games section... although I think now the firewall blocks most of them."

"Leaderboards... it's not... in our school... I don't think it happens in very many schools nowadays."

Survey engagement responses suggested stable/increased engagement among pupils using Eedi, but free-text responses underscored variation by format (e.g., some pupils resisted videos/practice questions).

Interviewed teachers described using performance overviews via displaying progress to motivate students and drive competition and setting clear quiz targets/using monitoring to prompt more practice. This resonates with the survey's reports of teachers using data to review misconceptions (80%), inform planning (77%), identify individual misconceptions (77%) and set quizzes (93%).

"For me it was trying to get them to complete more quizzes... if they got in that feedback cycle, then they weren't really pushing on."

"My class like the games section, so I'm like, 'Right, you need to complete this amount of quizzes and then you can go on that.'"

Convergence Coding Matrix

The matrix below synthesises RCT, survey, and interview evidence in a joint display organised around three themes: Set-Based Engagement & Progression, Connectivity, Accessibility and Culture, and Incentives and Motivational Features. For each construct we indicate whether strands Converge (C), Complement (Comp), Diverge (D), or are Silent (S).

Read-across from the matrix suggests that outcome gains are most plausible when access enables regular use and tasks are well-matched to pupil readiness, allowing teachers to embed retrieval-oriented routines and monitoring. Conversely, where device/home access is constrained, homework culture is weak, departmental capacity is stretched, or progression overshoots for middle/lower sets, engagement drops, and as a result, the expected “dose” may be suppressed. This could explain the attrition observed in the sample, non-significant ITT and observed heterogeneity. Motivational levers and platform features can boost throughput but are context-dependent (e.g., firewalls, school culture), which may further help to explain variation alongside the quantitative signals.

Major Theme 1: Set-Based Engagement & Progression				
<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Impact	ITT +ve, statistically significant ($d \approx 0.25$, 95% CI: 0.07 to 0.43). Subgroup exploration: non-PP +ve, PP +ve. Sensitivity with lower pre-test pupils +ve.	Platform adoption high; engagement mostly stable/increasing.	Works better for top sets; mid/lower sets stall when diagnostics escalate and in diagnose→resolve loops.	Comp (small average + heterogeneity aligns with “works for some cohorts/contexts”).
Progression pitch & feedback-cycle friction	No direct metric	Free-text flags “jumps” in diagnostic difficulty; some resistance to videos/practice format.	Two mechanisms: fast ramp into problem-solving; pupils “stuck” after errors. Demotivating for some.	C (strong cross-strand agreement on mechanism limiting progress for lower/middle sets).
Content mapping & depth	N/A	Mis-match with NI/Scottish Applications of Maths; requests for more lower-ability material and >5 items per set.	Teachers report limited lower-ability quizzes; desire Scottish set; 4th/5th diagnostic item “jumps”.	C (clear alignment on curriculum fit gaps).
Quiz configuration	N/A	Requests to assign >5 questions.	Teachers want longer sets for older cohorts; current cap constrains use.	C (surveys and interviews align).

Major Theme 2: Connectivity, Accessibility and Culture				
<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Access/IT constraints (devices/Wi-Fi/firewalls)	N/A	Top barriers: in-school devices/connectivity 43%; home access 40%.	From three laptop trolleys → one; booking pressure; old/broken devices; games/firewalls block; reliance on mobiles.	C (constraints plausibly depress the dose linked to gains).
Homework culture & home support	N/A	Teachers tie engagement to parental support/homework culture.	Weak homework culture (e.g., 60–65% completion in top set); limited consequences for non-completion.	C (shared explanation for lower homework uptake)
Department capacity & culture	N/A	Survey responses identified number of teachers not hitting dosage targets, described frequencies of usage practices low	Staffing pressure → Eedi as “another task”; informal championing (reminders/posters); planned alignment not resourced → usage drops to fortnightly.	Comp (explains organisational mechanism underlying dose shortfall which was identified via surveys).
Onboarding, training & communications	N/A	67% attended training; set-up largely straightforward.	Prior familiarity helped; some emails missed/bounced.	Comp (generally workable with communication gaps).

Major Theme 3: Incentives and Motivational Features				
<i>Construct</i>	<i>RCT signal</i>	<i>Survey finding</i>	<i>Interview insight</i>	<i>Convergence</i>
Voucher incentive (£30/month)	No direct metric; some students not meeting dosage.	Engagement often varied.	Voucher “massively increased” throughput; more quizzes completed per lesson/year group when publicised, had to try various approaches to sustain engagement.	Comp (teacher experiences of engagement reflect mixed dosage findings).
Gamification (avatars, games, leaderboards)	N/A	Mixed: format preferences vary; some pupils resist videos/practice.	Avatars blocked; games popular but often firewall-blocked; leaderboards culturally variable.	Comp (surveys show variation; interviews provide reasons or conditions).
Monitoring & goal-setting in class	N/A	Widespread use of data: misconceptions (80%), planning (77%), identify individuals (77%), set quizzes (93%).	Teachers display progress, set clear quiz targets, and use competition to prompt retrieval.	C (practice patterns align with dose–response).

Table 27. Convergence coding matrix

Discussion

Outcome and process data indicate that *Eedi's* strongest potential lies in sustained, routine use that embeds retrieval and diagnostic practice as part of everyday teaching. The average impact of the trial was small and non-significant in the first year, but grew to a moderate, statistically significant effect by 18 months ($d = 0.46$) and remained significant at 24 months ($d = 0.30$). Clear dose-response relationships within the intervention arm show that pupils who engaged more regularly made meaningfully greater progress over time. Real-world constraints around access, progression pitch, and workload moderated the level of engagement many schools could sustain, helping to explain why average effects took time to emerge and why they varied between schools. This pattern mirrors wider evidence on digital mathematics tools, which shows that learning benefits are achievable when teachers use platforms consistently and act on diagnostic data (Hillmayr et al., 2020; Witherspoon et al., 2024). The findings here therefore confirm that *Eedi's* impact is implementation-dependent, with measurable gains emerging where the platform is well-embedded and supported in practice.

Pupil Outcomes

Longitudinal ITT analyses showed small effects at 6 months ($d = 0.06$) and 12 months ($d = 0.17$), increasing to a moderate effect at 18 months ($d = 0.46$). By 24 months, the effect remained positive ($d = 0.30$). This trajectory suggests that benefits strengthened over time, peaking around 18 months before stabilising. The effects at 6 and 12 months were not statistically significant, but the effects at 18 months ($p = 0.0002$) and 24 months ($p = 0.0037$) were, with confidence intervals excluding zero. Within the intervention arm, usage patterns were consistently associated with attainment. These internal dose-response relationships align with wider evidence showing that regular independent practice combined with timely feedback drives learning gains (Hillmayr et al., 2020; Witherspoon et al., 2024).

The pattern of delayed onset of impact is noteworthy. Rather than indicating a lack of effect, the findings suggest that *Eedi* confers cumulative benefits as teachers and students integrate its use into their routines across multiple terms. This is plausible given the nature of the intervention: time was needed for teachers to embed diagnostic feedback and for students to adapt their study habits. The statistically significant results from the CACE analyses reinforce that the platform delivers durable impact when engagement is high. Analogous findings appear in other research. A recent multi-state RCT of ASSISTments also found no significant test score difference after one year of use in Grade 7, but a statistically significant improvement (Hedges' g

≈ 0.10) became evident one year later (Feng et al., 2023). This parallel strengthens confidence in the conclusion that sustained implementation is key.

For Pupil Premium (PP) students, effect sizes were small and non-significant at 6 months ($g = 0.111$, 95% CI: -0.101, 0.323, $p = 0.290$) and 12 months ($g = 0.142$, 95% CI: -0.065, 0.348, $p = 0.168$), before increasing to a moderate and statistically significant effect at 18 months ($g = 0.370$, 95% CI: 0.106, 0.635, $p = 0.0075$), then decreasing and becoming non-significant again at 24 months ($g = 0.097$, 95% CI: -0.144, 0.338, $p = 0.418$). For non-PP students, effects were generally larger, starting small at 6 months ($g = 0.034$, 95% CI: -0.151, 0.218, $p = 0.709$), increasing at 12 months ($g = 0.171$, 95% CI: -0.018, 0.360, $p = 0.074$), and reaching their peak at 18 months ($g = 0.508$, 95% CI: 0.265, 0.751, $p = 0.0002$) and remaining substantial at 24 months ($g = 0.400$, 95% CI: 0.171, 0.629, $p = 0.0012$). These patterns suggest that the intervention may have a delayed and more pronounced impact for non-PP students, while effects for PP students appear to peak around 18 months before diminishing. The direction and timing of these effects reinforce the process evaluation findings that dosage and scaffolding are especially important for pupils starting at lower attainment levels.

CACE analysis provided the clearest evidence of impact: higher compliance consistently produced larger effects, with most estimates statistically significant from 12 months onwards. For example, at 12 months, pupils meeting the highest compliance threshold achieved $d = 1.04$ (95% CI: 0.94 to 1.15), rising to $d = 1.45$ (95% CI: 1.29 to 1.60) at 18 months. These findings highlight that *Eedi's* effectiveness depends on sustained, high-fidelity implementation: when pupils and teachers engage regularly and consistently, the platform produces educationally meaningful improvements that accumulate over time.

Progression, differentiation and the “diagnose → resolve” loop

One of the key advantages of digital homework platforms lies in their potential to differentiate support according to pupil need. *Eedi's* diagnostic questioning, scaffolded feedback, and challenge items are designed to achieve this. However, the process data suggest that differentiation was only partially realised in practice. Lower-attaining pupils were sometimes drawn into repetitive diagnose-and-resolve loops, while higher-attaining pupils tended to progress more smoothly through quizzes. Teachers also noted curriculum-mapping gaps and requested greater flexibility in assignment length to tailor work to different cohorts.

These implementation patterns help explain why the trial's average impact took time to emerge, and they contextualise the moderate gains among lower-attaining pupils, which were statistically significant at 12, 18 and 24 months. The evidence indicates that where differentiation mechanisms were well aligned with pupil readiness, and where access allowed regular engagement, pupils made measurable progress, consistent with the significant dose-response relationships seen. This aligns with research on diagnostic questioning and the role of confidence and misconception handling in secondary mathematics (Foster et al., 2022). Enhancing scaffolding and curriculum alignment could therefore help strengthen these effects further, particularly for pupils starting from lower attainment levels or those facing disadvantage.

Fidelity and Teacher Workflow

Rapid progression accounted for varied pupil engagement, with some students requiring more support than others. Disengaged usage is a known challenge in online learning, often linked to access or motivation barriers. IPE surveys highlighted the importance of school IT reliability and home access. These constraints remain key moderators of dosage. Moreover, school-based barriers also limited usage. Teachers reported workload benefits, especially reductions in marking, but some practical barriers (time, curricula alignment, multiple platforms in use) limited uptake. Average usage of the platform was below intended levels, which likely diluted overall ITT effects. However, the compliance analysis shows that those teachers and pupils who did use *Eedi* consistently achieved significantly larger gains.

Implications

The synthesis of findings points to several actionable implications:

- **Highlight value-added:** Teachers consistently reported reduced marking loads and more efficient progress monitoring. Many used *Eedi* data to review misconceptions, plan lessons, and target support. These tangible workload and insight benefits should be foregrounded to drive adoption and sustained use.
- **Expand onboarding:** Limited feature uptake reflected low awareness rather than resistance. Streamlined onboarding and "quick start" guidance can accelerate adoption.
- **Enhance differentiation and support for struggling learners:** Exploratory analyses showed that pupils with lower baseline scores made substantial gains,

statistically significant at 12, 18 and 24 months. Refining progression pitch, feedback loops, and scaffolds will help ensure that differentiation mechanisms work reliably across all attainment levels.

- **Leverage data for teacher intervention:** Diagnostic analytics remain valuable but underused. Simplifying data dashboards and automating actionable insights could help teachers intervene earlier and more consistently, translating diagnostic value into learning gains.
- **Continue rigorous evaluation and iteration:** Average ITT effects were positive throughout and statistically significant at 18 and 24 months, and CACE and usage analyses revealed larger improvements still among highly adherent users. This highlights that context, fidelity, and engagement remain critical drivers of impact. *Eedi's* ongoing school test-bed and iteration model provides an ideal framework to continue refining implementation strategies and testing whether these strengthen group-level effects further.

Conclusion

Findings from this two-year randomised controlled trial evaluation indicate that *Eedi* provides a robust foundation for improving mathematics learning when it is used regularly and as intended. Across the 24-month period, the average intervention effects (ITT) were positive at every time point, increasing from small at six months to a moderate, statistically significant effect by 18 months ($d = 0.46$) and remaining significant at 24 months ($d = 0.30$). The consistent upward trajectory suggests that *Eedi's* potential benefit strengthens over time with continued use.

Subgroup analyses showed positive effects for both Pupil Premium and non-Pupil Premium pupils: the effect for Pupil Premium pupils was statistically significant at 18 months, while non-Pupil Premium pupils showed significant effects at 18 and 24 months. Sensitivity analyses focusing on pupils with lower baseline attainment revealed a similar pattern: these pupils demonstrated large improvements, peaking at 18 months and statistically significant at 12, 18 and 24 months. This pattern suggests that *Eedi* may be particularly beneficial for pupils who begin with lower attainment levels. The CACE (Complier Average Causal Effect) analyses further illuminate this pattern. When adherence was high, effect sizes increased markedly, reaching statistical significance at almost every compliance threshold from 12 months onwards. This confirms that the intervention's impact is strongly implementation-dependent: pupils and schools that maintained consistent engagement achieved meaningfully greater gains.

Teacher surveys and interviews add qualitative depth to these results. Teachers reported that *Eedi* reduced marking and planning workload, supported formative assessment, and offered actionable insights into misconceptions. Barriers to impact were primarily practical: limited device access, weak homework culture, and inconsistent embedding across departments. These factors help explain why the average ITT effects were modest even though high-engagement users demonstrated clear progress.

Together, both the RCT and IPE evidence converge on a consistent message. *Eedi's* impact is durable, scalable, and achievable in real classrooms when conditions for regular use are met. The data show that sustained engagement over multiple terms translates into measurable learning gains, particularly for lower-attaining pupils and those who adhere closely to the intended model of use. Identified challenges are tractable and can be addressed through refined onboarding, scaffolding, and teacher support.

In summary, this evaluation provides strong evidence that *Eedi* is a promising digital platform for supporting Key Stage 3 mathematics learning. Average trial-level effects were positive throughout and statistically significant at 18 and 24 months, and significant adherence and usage effects show that impact grows with sustained, consistent use. Continued iteration focused on access equity, sustained weekly retrieval, and targeted scaffolding is likely to unlock stronger, statistically robust gains at scale.

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>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Di Pietro, G., & Muñoz, J. C. (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

Foster, C., Woodhead, S., Barton, C., & Clark-Wilson, A. (2022). School students' confidence when answering diagnostic questions online. *Educational Studies in Mathematics*, 109(3), 491–521. <https://doi.org/10.1007/s10649-021-10084-7>

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., Feters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *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*. Pearson.

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>

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>

- 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/10.3386/w27476>
- 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
- UNICEF. (2025). UNICEF digital education strategy 2025–2030. UNICEF.
- Vanacore, K. (2025). Attrition analysis report: Eedi RCT. Internal technical report.
- 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>
- Witherspoon, E. B., Pardo, M., & Walters, K. (2024). Teacher use of an online platform to support independent practice in middle school mathematics during COVID-19 disruptions. *AERA Open*, 10. <https://doi.org/10.1177/23328584241230054>



I = Eedi

Contact

support@whatworked.education

**WhatWorked Education
Harrison House
Hawthorn Terrace
Durham
DH1 4EL**