EdTech Hub

Clear evidence, better decisions, more learning.

Randomised Controlled Trial Protocol

Digital personalised learning to improve literacy and numeracy outcomes in early grade Kenyan classrooms

Date October 2023

Authors Louis Major
Rebecca Daltry
Mary Otieno
Kevin Otieno
Annette Zhao
Jessica Hinks
Chen Sun
Aidan Friedburg

DOI 10.53832/edtechhub.0176

UKaid from the British people

THE WORLD BANK

unicef for every child

#EdTechHub @GlobalEdTechHub edtechhub.org

About this document

Recommended citation

Major, L., Daltry, R., Otieno, M., Otieno, K., Zhao, A., Hinks, J., Sun, C., & Friedburg, A. (2023). *Randomised Controlled Trial Protocol: Digital personalised learning to improve literacy and numeracy outcomes in early grade Kenyan classrooms*. [Methodology publication]. EdTech Hub. https://doi.org/10.53832/edtechhub.0176. Available at https://docs.edtechhub.org/lib/JNH4277Z. Available under Creative Commons Attribution 4.0 International.

Reviewers

Ricardo Sabates

About EdTech Hub

EdTech Hub is a global research partnership. Our goal is to empower people by giving them the evidence they need to make decisions about technology in education. Our evidence library is a repository of our latest research, findings and wider literature on EdTech. As a global partnership, we seek to make our evidence available and accessible to those who are looking for EdTech solutions worldwide.

EdTech Hub is supported by UKAid, Bill & Melinda Gates Foundation, World Bank, and UNICEF. The views in this document do not necessarily reflect the views of these organisations.

To find out more about us, go to edtechhub.org/. Our evidence library can be found at docs.edtechhub.org/lib/.

Contents

List of figures and tables *4*
Abbreviations and acronyms *5*
1. Background to the study **6**
1.1 Investigators 6
1.2 Research objective 7
1.3 Overarching research design and timeline 7
1.4 Intervention and operational context 8
1.5 Rationale for an RCT 9
1.6 Purpose of this document 11
2. Methodological approach **13**
2.1 Treatment and control groups 13
2.2 Sample selection 15
2.3 Randomisation strategy 19
2.4 Learning assessment 21
2.5 Supplementary questionnaires 25
2.6 Integration with EIDU app data 26
2.7 Analysis strategy 27
3. Methodological limitations **34**
3.1 Contextual relevance of RCT conditions 34
3.2 School-level randomisation 34
3.3 Learner-level randomisation 35
3.4 Phased rollout of treatment arm 36
3.5 Participant engagement in the research 37
3.7 Logistical challenges of assessment 37
4. Ethical considerations **39**
4.1 The ethics of randomised controlled trials 39
4.2 Informed consent 39
4.3 Safeguarding 40
4.4 Anonymity, confidentiality, and privacy 40
4.5 Data management 41
4.6 Ethics Review Committee 41
References **43**
Annex A: full research timeline **50**
Annex B: full range of a priori power calculations **51**

Figures and tables

Figure 1. Allocation of new ECDOs within the sub-county structure, to prevent contamination between the treatment and control group schools 20

Table 1. Primary research team for the RCT 6
Table 2. Overview of the methodological approach 13
Table 3. Staggered rollout of the treatment arm 14
Table 4. Overview of power calculations to determine minimum sample size 15
Table 5. Distribution of school sample (n = 375) across sub-county strata 17
Table 6. 'Before' and 'after' comparison of school distribution across sub-county strata, following exclusions 18
Table 7. Thresholds for final sample of treatment and control schools, by sub-county 18
Table 8. Timeline of RCT assessment rounds 22
Table 9. IDELA assessment items 23
Table 10. Overview of supplementary characteristics data 25
Table 11. Full research timeline 50
Table 12. Full set of power calculations 51

Abbreviations and acronyms

CBC	Competency-Based Curriculum
DPL	Digital personalised learning
DPL-SP	Digital personalised learning and structured pedagogy
ECDOs	Early Childhood Development Officers
ICC	Intra-class correlation
IDELA	The International Development and Early Learning Assessment
KICD	Kenya Institute of Curriculum Development
KEMRI	Kenya Medical Research Institute
LAYS	Learning-adjusted years of schooling
MELQO	Measuring Early Learning Quality and Outcomes
NACOSTI	National Commission for Science, Technology and Innovation
PP1 / PP2	Pre-primary grades 1 and 2
RCT	Randomised controlled trial
SEND	Special educational needs and disability
SP	Structured pedagogy
WERK	Women Educational Researchers of Kenya

1. Background to the study

As part of its portfolio of research studies,[1] EdTech Hub launched a study in 2022 to rigorously investigate how a classroom-integrated, digital personalised learning (DPL) tool can most effectively support early grade numeracy and literacy outcomes in Kenya. This multi-strand study is planned to run until 2025 and explores a DPL tool developed by EIDU — an education technology developer, providing a low-cost DPL platform to pre-primary and primary learners in low- and middle-income settings.

1.1 Investigators

The randomised controlled trial (RCT) research team was assembled through EdTech Hub (see Table 1), in partnership with Women Educational Researchers of Kenya (WERK), a professional association of researchers in education and social science.[2] Additionally, WERK is responsible for the recruitment, training, and coordination of a team of 42 enumerators for all assessment rounds.

Table 1. *Primary research team for the RCT*

Research partner	Name	Associate research organisation	Role and responsibilities
EdTech Hub	Dr Louis Major	Senior Lecturer, University of Manchester	Principal investigator
	Rebecca Daltry	Jigsaw	Research manager
	Dr Annette Zhao	Jigsaw	Research associate (quantitative analysis)
	Jessica Hinks	Jigsaw	Research associate
	Dr Chen Sun	University of Manchester	Quantitative analysis advisor
Women Educational Researchers of Kenya	Dr Mary Otieno	Senior Lecturer, Kenyatta University	Senior researcher
	Kevin Otieno	Kindergarten Experts International	Research assistant

[1] https://edtechhub.org/evidence/edtech-hub-research-portfolio/ Retrieved 17 October 2023
[2] https://werk.co.ke/ Retrieved 17 October 2023

1.2 Research objective

The multi-strand research study is framed by the overarching research question:

How can a classroom-integrated, digital personalised learning tool (EIDU) most effectively support early grade numeracy and literacy outcomes in Kenya?

Each strand of the research is also informed by tailored research questions. The question relevant to the 'learning outcomes' strand, which involves the cluster RCT outlined in this protocol, is:

What is the impact of classroom-integrated digital personalised learning on early grade numeracy and literacy learning outcomes?

1.3 Overarching research design and timeline

The larger study as a whole consists of multiple research strands:

- **A pedagogical strand**, exploring the pedagogical implications of integrating DPL into pre-primary classrooms through a design-based research approach, integrating mixed-methods research and innovation strategies.
- **A learning outcomes strand**, to assess the impact of classroom-integrated DPL on numeracy and literacy outcomes.
- **An equality strand**, to identify effective approaches for learner selection that might promote equal device usage through the introduction of iterative innovation interventions.
- **An adaptivity and data feedback strand**, to assess the potential of optimising software affordances and integrating digital assessment tools into the teaching and learning process.

As part of the 'learning outcomes' strand, this protocol outlines the methodological strategy for a cluster RCT to assess the impact of a classroom-integrated DPL tool on numeracy and literacy outcomes in Kenyan pre-primary grades. Baseline assessment for the RCT took place in October 2022 (the start of Term 3), midline assessment in May 2023 (the start of Term 2), and endline assessment is scheduled for October 2023 (the end of Term 3). A full research timeline is available in Annex A.

1.4 Intervention and operational context

1.4.1 EIDU's classroom-integrated approach: aligning DPL with structured pedagogy

EIDU's classroom-integrated implementation approach combines a DPL tool with structured pedagogy (SP) — henceforth called the EIDU 'DPL-SP' model.

The central, digitised component of the EIDU DPL-SP model is the EIDU application (app), installed on low-cost Android devices. This is built by EIDU software engineers based in Germany and Kenya. The app is content-agonistic — it has been developed to generate data-driven insights into learning outcomes and usage from content plugged into the software. The Android devices onto which the app is installed are provided by EIDU but paid for (at least in part) by county governments.

The EIDU app provides access to two types of content: a learner-facing DPL tool and teacher-facing lesson plans from the Tayari structured pedagogy (SP) programme. The DPL tool provides an interface with digitised formative and summative assessment units, personalised to learners' individual history (↑Friedberg et al., 2023). Content focuses on numeracy and literacy, consisting of digital formative assessment units open-sourced by onebillion[3] (a non-profit organisation), and assessment exercises developed by EIDU. The content units, approved by the Kenya Institute of Curriculum Development (KICD), are structured in strands and sub-strands to match the Kenyan competency-based curriculum (CBC). Since learners have individual EIDU profiles, the adaptive nature of the EIDU app calculates the most relevant learning unit within each sub-strand for the individual, according to their learning history on the app.

The DPL tool complements the second functionality of the EIDU app: the provision of the Tayari lesson plans. The Tayari SP programme was developed by RTI[4] (an independent non-profit research institute) and the Kenyan government for pre-primary grades in Kenya and has been the focus of multiple evaluations (e.g., ↑Ngware et al., 2018; ↑Piper et al., 2018b). The programme features training and ongoing support from early childhood development officers (ECDOs). EIDU have digitised the lesson plan component of Tayari SP, meaning that numeracy and literacy content (covering a 27-week period, 5 days per week) is available on the EIDU app. Since the Tayari lesson plans have been mapped to Kenya's CBC in the same way as the DPL exercise units, when a teacher marks a lesson as

[3] https://onebillion.org/ Retrieved 17 October 2023
[4] https://www.rti.org/ Retrieved 17 October 2023

'complete' on the app, the app can then present the learning unit(s) most relevant to that's day's teaching when the DPL tool is handed to learners.

As such, the EIDU app provides a DPL model which is integrated within, rather than supplementary to, classroom practice. The mapping of DPL exercise units and Tayari lesson plans to the CBC means teaching and learning are aligned both on and off the EIDU platform. Limited research has, to date, evaluated a DPL model aligned to curriculum and classroom practice in this manner (↑Major et al., 2021).

1.4.2 The structure of pre-primary education in Kenya

In Kenya, pre-primary education is governed by the National Pre-Primary Education Policy of 2017 (↑Republic of Kenya Ministry of Education, 2017) and the National Pre-Primary Education Policy Standard Guidelines of 2018 (↑Republic of Kenya Ministry of Education, 2018). The 2017 policy is operationalised through the standards to help "ensure that quality services are delivered efficiently and effectively at all times in all pre-primary education centres/institutions in Kenya" (↑Republic of Kenya Ministry of Education, 2018, p.iv).

Kenya's pre-primary education has a two-year structure: PP1 (4–5-year-olds) and PP2 (5–6-year-olds). The 2017 policy outlines how the responsibility for the delivery of this service is devolved to the county governments. The governments and their officers are responsible for elements of pre-primary education (PPE) delivery, including putting in place overarching management and governance structures, developing strategies and budgets for resource management, and ensuring quality service delivery.

Across all counties, groups of approximately 20 schools are currently assigned to a county ECDO — although, in practice, this can range from 15 to 75 schools per ECDO. These schools are clustered geographically close together in 'zones'. An ECDO regularly visits schools within their grouping (once a month to once a term) to check in and provide guidance to ensure the quality of services.

1.5 Rationale for an RCT

RCTs have been increasingly used to investigate the educational effects of DPL (↑Major et al., 2021). In an RCT of onebillion's mathematics content in Malawi, significant learning gains in numeracy skills were observed, compared to standard mathematics instruction or using tablet devices without the software (↑Pitchford, 2015). A study to measure the impact of onebillion's content on learning outcomes also found that, when implemented at the start of primary education (before significant gender discrepancies become established), use of the app prevented significant gender gaps emerging in mathematics (↑Pitchford et al., 2019). Similarly, an RCT of the Tayari SP programme (focused primarily on teacher training and classroom instruction) in Kenyan pre-primary schools found a standardised effect size of 0.33 on learner achievement, with 62% of schools in treatment groups scoring higher than those in control groups (↑Ngware et al., 2018). The present RCT will complement and expand upon these existing studies and others (e.g.,↑Piper et al., 2018b). While individual components of the EIDU DPL-SP model (including the onebillion learning content and Tayari SP programme) have been analysed, a classroom-integrated model which aligns a DPL tool with SP has not been rigorously evaluated in the Kenyan context.

While claims of RCTs representing a 'gold standard' of educational research are increasingly considered to be exaggerated (e.g., ↑Engeström, 2011; ↑Sims et al., 2023), and a number of large-scale RCTs have been criticised for potentially being uninformative (↑Lortie-Forgues & Inglis, 2019), the education RCT remains a powerful research tool for assessing interventions designed to help children learn (↑Styles & Torgerson, 2018). This is because, when rigorously undertaken, RCTs can provide robust evidence of causal inference and quantified impact, have the potential to identify educational initiatives that "might do more harm than good", and identify areas where investment of often limited resources might be accelerated (↑Torgerson & Torgerson, 2012). Providing caution against overinterpretation is exercised, opportunities for subgroup analyses to establish "what works for whom" are also possible. A example would be analysing the effects of an intervention not just with regard to the sample as a whole, but also in relation to whether the effects differ between subgroups of students, such as girls compared to boys or those from differing socio-economic backgrounds (↑Connolly et al., 2018).

The present RCT is designed in alignment with the other research strands outlined in Section 1.3. It builds on two cycles of rigorous design-based research to understand the pedagogical context of the DPL-SP model. Conducting an RCT in the context of this strand of research, among others, therefore seeks to mitigate the well-documented limitations of RCTs. It also seeks to address the requirement that a new educational intervention has to be designed and thoroughly understood before it can be more extensively tested. The pedagogical strand of research can investigate 'how' and 'why' a classroom-integrated DPL model can be implemented in Kenyan pre-primary settings, in contrast with the RCT strategy, which primarily considers whether the intervention impacts learning outcomes.

Important ethical issues associated with education RCTs, including threats to internal and external validity, as well as potential methodological limitations such as differential drop-out and random assignment, are outlined later.

1.6 Purpose of this document

This protocol is being disseminated prior to endline data analysis to promote transparency and a comprehensive understanding of the RCT approach and design. The final analytical strategy was determined after cleaning and sorting baseline and midline data. The protocol has been iteratively developed following external expert input and review to allow constructive modifications at appropriate stages (↑Kendall, 2003): the protocol was first drafted and externally reviewed in July / August 2022; contextual factors encountered during baseline assessment were accounted for in November / December 2022; methodological decisions regarding midline assessment were integrated and externally reviewed in April / May 2023; and this final version, including methodological decisions regarding endline assessment and the final analytical strategy, was finalised and externally reviewed in September / October 2023, ahead of endline data analysis. This strategy was necessary because there was a lack of preliminary data to inform the research plan, and as conducting a pilot quantitative study beforehand was not feasible.

↑Coskinas et al., (2020) outline how ongoing RCTs may require changes based on new information. Reactive revisions, potentially informed by foundational research revealing new insights, can arise due to seeking greater statistical power and/or to accommodate unexpected characteristics of preliminary data. Discretionary decisions allow flexibility relating to analysis details. Examples include whether to exclude a small proportion of participants found to have been ineligible after randomisation (based on pre-randomisation factors), refinement of an ambiguous endpoint definition, whether to adjust for a particular baseline covariate or the choice of statistical test to compare randomised groups on a given endpoint.

Both reactive revisions and discretionary decisions are mechanisms that can lead to changes in the design or analysis of an ongoing RCT, and such changes may bias results. This potential for bias depends on whether a change is based on information completely independent of treatment allocation or if it is related to treatment allocation (↑Coskinas et al., 2020). Changes should be regarded as legitimate if they are based on information independent of treatment allocation and the revision is otherwise sound (for example, if the revised question is still regarded as important, the revised design has adequate power, the study remains ethical, etc). Examples outlined in this protocol include decisions to include or exclude schools in the final sample due to contextual factors (for instance, excluding schools after baseline assessment which were found not to have the required number of learners to achieve adequate power, due to inaccurate school enrolment reporting) or ethical factors (e.g., replacing a treatment school with another randomly allocated treatment school prior to baseline assessment, due to ethical concerns around the well-being of one teacher). Such legitimate changes should not be seen as threats to the credibility of a well-conducted RCT and such studies remain valid (↑Coskinas et al., 2020), providing changes have not been made post hoc to manufacture positive results (↑Orkin et al., 2021).

In sharing this detailed RCT protocol in full, the intention is to mitigate any potential risks to trial integrity by demonstrating the process of systematic documentation implemented. This intends to facilitate a fair appraisal of results and to demonstrate that selective reporting has not taken place.

2. Methodological approach

The evaluation is designed as a stratified, two-arm, cluster-randomised controlled trial. It involves over 3,000 learners from 297 early childhood development centres across four sub-counties of Murang'a county (Gatanga, Maragua, Kandara, and Mathioya). The primary outcome measure is the assessment of emergent numeracy and literacy domains using the International Development and Early Learning Assessment (IDELA) assessment battery.[5] More details of each of these methodological components, including the sampling, randomisation, assessment, and analysis strategy, are outlined in Table 2 below.

Table 2. *Overview of the methodological approach*

Number of arms	Two-arm (treatment and control)
Trial type	Stratified and cluster-randomised
Unit of randomisation	School
Stratification variables	Sub-county
Assessment rounds	Baseline, midline, and endline
Primary outcome — variable	Emergent numeracy and literacy outcomes
Primary outcome — measure	IDELA assessment battery (emergent numeracy and literacy domains, with two substituted Measuring Early Learning Quality and Outcomes (MELQO) assessment battery literacy items)

2.1 Treatment and control groups

The RCT involves drawing comparisons in terms of learning gains between one treatment and one control group.

The treatment arm consists of schools that received the EIDU DPL-SP model. Schools received the treatment in a staggered approach (see Table 3 below). The rollout first consisted of learners receiving the DPL tool on one Android device, with teachers receiving associated training. Following one term of implementation, teachers received training on the Tayari SP programme from the county's ECDOs (who themselves had received training from EIDU on the SP programme) and gained access to the digitised Tayari lesson plans on the EIDU app. As such, the combined

[5] https://idela-network.org/ Retrieved 17 October 2023

DPL-SP model, including ongoing support from the ECDOs during term time, was operational over the full 2023 school year, after one initial term of the DPL tool only in 2022. This staggered approach is implemented by EIDU as a way to integrate the model into classrooms without overwhelming teachers and learners with the full DPL-SP from the outset.

Table 3. *Staggered rollout of the treatment arm*

Date	Time in relation to the Kenyan school year	Rollout stage of the treatment arm
28 September 2022	The start of Term 3	Town Hall meetings (for teachers to be introduced to EIDU, be given low-cost Android device with pre-installed EIDU app and receive training on app use).
w/c 26 September 2022	The start of Term 3	Learners begin using the DPL tool on the EIDU app.
w/c 9 January 2023	Prior to Term 1	Tayari teacher training for the SP programme (including using the digitised lesson plans).
w/c 23 January 2023	The start of Term 1	Full DPL-SP model begins in classrooms: second Android device given to schools and EIDU app updated so that teachers are able to access Tayari lesson plans and learners are able to access the DPL tool.

The control arm consists of schools who are yet to receive the EIDU DPL-SP model. Teachers in these schools continue to deliver the non-digitised CBC curriculum to PP1 and PP2 learners and receive the same level of support from ECDOs as prior to the RCT. None of the non-digitised components of the Tayari SP programme are received by the control schools during the period of the RCT. Instead, the control schools are due to receive the EIDU DPL-SP model, with phased rollouts due to begin in January 2024.

It is a recognised critique of education interventions that providing additional inputs into the classroom (including technology), without addressing the pedagogical implications of their uptake by teachers and learners, is rarely an effective approach (↑Banerjee et al., 2023; ↑Global Education Monitoring Report Team, 2023). For this reason, this RCT is one component of a broader, multi-strand research study, which is exploring other elements of the pedagogical integration of EIDU's DPL-SP model into Kenyan pre-primary classrooms. This involves a collaborative partnership between teachers and researchers, to investigate the integration of EIDU's DPL-SP model into classroom practice through a design-based research approach. The aim of this RCT is not to 'prove' the efficacy of the EIDU tool in isolation, but to recognise and interpret the findings in relation to further rigorous mixed-methods research.

2.2 Sample selection

This section outlines the process by which the sample was narrowed down from a possible population of all government pre-primary schools across three Kenyan counties to 292 schools in four sub-counties of Murang'a county. All methodological decisions taken were informed through close coordination and consultation between the combined EdTechHub and WERK research team, research advisors (both internal and external to EdTech Hub), and EIDU, who have significant operational experience. This strategy helped to ensure pragmatic decision-making that balanced feasibility (e.g., operational factors) and research-related decisions.

2.2.1 Power calculations

A range of power calculations was made in May 2022 (available in Annex B), using the WebPower package in R (wp.crt2arm command; ↑McNulty, 2021, p. 11; ↑Zhang, 2022) to estimate the minimum sample size required. A final minimum sample size of 276 schools (138 schools in the treatment and control groups) was selected to detect a mean effect size of 0.2 standard deviations with an assumed power of 80%, building on the work of ↑Piper et al. (2018b). This sample size accounts for an estimated attrition rate of up to 20% and uses an intraclass correlation (ICC) of 0.2. It allows for 10 randomly selected learners to be assessed per school, which was deemed feasible with the selected assessment strategy.

Table 4. *Overview of power calculations to determine minimum sample size*

School cluster sample size (incl. 1 treatment arm & control)	Learner assessment sample size per school	Minimum detectable effect size	Power	Intraclass correlation (ICC)
220 (276 when accounting for 20% attrition)	10	0.2	0.8	0.2

2.2.2 Sample population

Over the course of 2022–23, EIDU has begun rolling out the DPL-SP model to government schools in three Kenyan counties (Embu, Murang'a, and Kiambu). Within this population, there are approximately 1,500 schools and 75,000 children, with an average of 30 children per class (ranging between 4 and 178 learners per grade in PP1 or PP2). The observable characteristics of the sub-counties vary, including in the rural / urban ratio and teacher–pupil ratio. Since this research is taking place in parallel to the rollout of the EIDU DPL-SP model in these three new counties, these comprised the potential population for the RCT.

Eligibility criteria were developed in May 2022 to determine the sub-counties (of the 24 sub-counties across Embu, Murang'a, and Kiambu) from which the randomised sample would be taken. This consisted of three stages. First, eight sub-counties were excluded on the basis of having skewed characteristics data (including the number of schools and teacher–pupil ratio) — this was calculated using data provided by county governments. This step was taken to avoid an especially heterogeneous sample, therefore potentially diluting any effect size observed. Second, a further nine sub-counties were excluded, taking into account the operational practicalities of conducting an RCT, to ensure the feasibility and efficiency of the RCT implementation.

Applying these criteria narrowed the potential sample to seven sub-counties:

- Manyatta and Runyenjes from Embu county
- Kahuro, Maragua, Kandara, Mathioya, and Gatanga from Murang'a county.

To select the final sample of sub-counties, data on all possible schools in each sub-county were compared, in addition to considering relevant census data from the 2019 Kenya Population and Housing Census (↑Kenya National Bureau of Statistics, 2019). A decision was taken to exclude schools with fewer than 40 learners in total, and fewer than 15 learners in PP1 and PP2 classes, to account for pupil absence on assessment days as well as possible school withdrawal. While this was a practical decision to enable the assessment of a minimum of 10 randomly selected learners per school, it was recognised that a mitigation strategy was required to account for the potential impact of selection bias caused by excluding schools with very small class sizes. The size of PP1 and PP2 classes at each school was therefore recorded at each assessment point, in order to gain insights as to whether classroom size correlates with learning outcomes and the impact of the treatment arm (see more in Section 2.5).

It was also decided to prioritise the inclusion of sub-counties which have the closest comparable socio-economic characteristics (urbanisation rate of the county, population size, school-attending population size, unemployment rate, and property) to ensure comparability and an appropriate sample size. This led to the exclusion of the two sub-counties from Embu County, considering the relatively smaller numbers of schools and pupils attending schools in comparison with Murang'a County.

Furthermore, Kahuro sub-county was excluded based on its relatively distinct socio-economic characteristics (low unemployment rate and high property ownership).

As such, the sub-counties of Maragua, Kandara, Mathioya, and Gatanga were selected as the final sample. All four sub-counties are within Murang'a County and include 375 government pre-primary schools. The distribution of these schools across the four sub-counties is presented in Table 5.

Table 5. *Distribution of school sample (n = 375) across sub-county strata*

Sub-county (in Murang'a county)	Number of schools	% of total potential sample
Maragua	n = 105	28%
Mathioya	n = 76	20%
Kandara	n = 88	24%
Gatanga	n = 106	28%

2.2.3 Sample size

Since a minimum of 276 schools was required to satisfy the power calculations, applying exclusion criteria further narrowed the sample from 375 to 316 schools. Following this, a process of random sorting and allocation of treatment and control took place.

The exclusion criteria to narrow the sample down to 316 schools included the following:

- The exclusion of schools for one or more practical reasons (n = 11):

  a. because two schools share the same building — which could cause contamination if the schools are allocated to two different groups (i.e., treatment and control) but occupy the same physical location;

  b. because of unresolved disputes between schools and the county government, which would mean their cooperation with the study could not be guaranteed across the whole year;

  c. if they are a specialist school for special educational needs and disability (SEND); currently the EIDU platform does not adequately address the needs of learners with SEND (although EIDU are actively developing approaches to address this).

- Exclusion of schools with fewer than 15 learners in either PP1 or PP2 (n = 48). Student data from the 2021/22 academic year was used as a basis for this estimate.

Following this selection of the possible 316 school sample, a descriptive 'before' and 'after' comparison was undertaken to investigate the impact of exclusion criteria on the distribution of schools across the four sub-counties. As outlined in Table 6, the criteria had limited impact.

Table 6. *'Before' and 'after' comparison of school distribution across sub-county strata, following exclusions*

	Pre exclusions		**Post exclusions**	
Sub-county (in Murang'a county)	Number of schools	% of total potential sample	**Number of schools**	**% of total potential sample**
Maragua	n = 105	28%	**n = 90**	**28%**
Mathioya	n = 76	20%	**n = 59**	**19%**
Kandara	n = 88	24%	**n = 76**	**24%**
Gatanga	n = 106	28%	**n = 91**	**29%**

Finally, the following 'thresholds' (outlined in Table 7) were set as the minimum number of treatment and control schools to be allocated across the four sub-counties. This was calculated ahead of randomisation, the process for which is outlined in the following section.

Table 7. *Thresholds for final sample of treatment and control schools, by sub-county*

Sub-county (in Murang'a county)	% of total potential sample (n = 276)	Minimum number of schools for RCT sample
Maragua	28%	78 (39 treatment, 39 control)
Mathioya	19%	52 (26 treatment, 26 control)
Kandara	24%	66 (33 treatment, 33 control)
Gatanga	29%	80 (40 treatment, 40 control)

2.3 Randomisation strategy

The randomisation strategy consisted of two stages:

1. allocation of schools to treatment and control groups, which involved stratified, cluster-randomisation;
2. random selection of learners to be assessed within each school, stratified by gender.

2.3.1 School-level randomisation (allocation of treatment and control groups)

The random allocation of schools to treatment or control groups took place at the school level, but within the stratified structure of the four sub-counties included in the sample. Considering the process through which ECDOs are allocated to schools across the county, the decision was made to retain the sub-counties as strata. As outlined above, groups of approximately 20 schools within each sub-county are assigned to an ECDO by the county government, with groups clustered geographically close together in 'zones'. It was thus not feasible to randomise across all four sub-counties, as this would not account for the zonal clustering of ECDOs. Furthermore, ECDOs in the treatment group took part in training to support the delivery of the Tayari SP programme, so there was a risk of contamination between the treatment and control groups.

The initial random allocation of schools to treatment and control groups took place in August 2022. First, schools in each sub-county sample (n = 316) were randomly sorted so the initial 276 could be selected. These 276 schools (and the remaining 40) were randomly assigned to 'treatment' or 'control'. Randomisation was affected in Microsoft Excel following guidance from the Poverty Action Lab (↑J-PAL, no date). Descriptive statistics (count, mean, and median), based on student data provided by the Murang'a county government from the 2021/22 academic year, were used to confirm that the student assignment to the treatment and control groups were comparable and not adversely influenced by outliers.

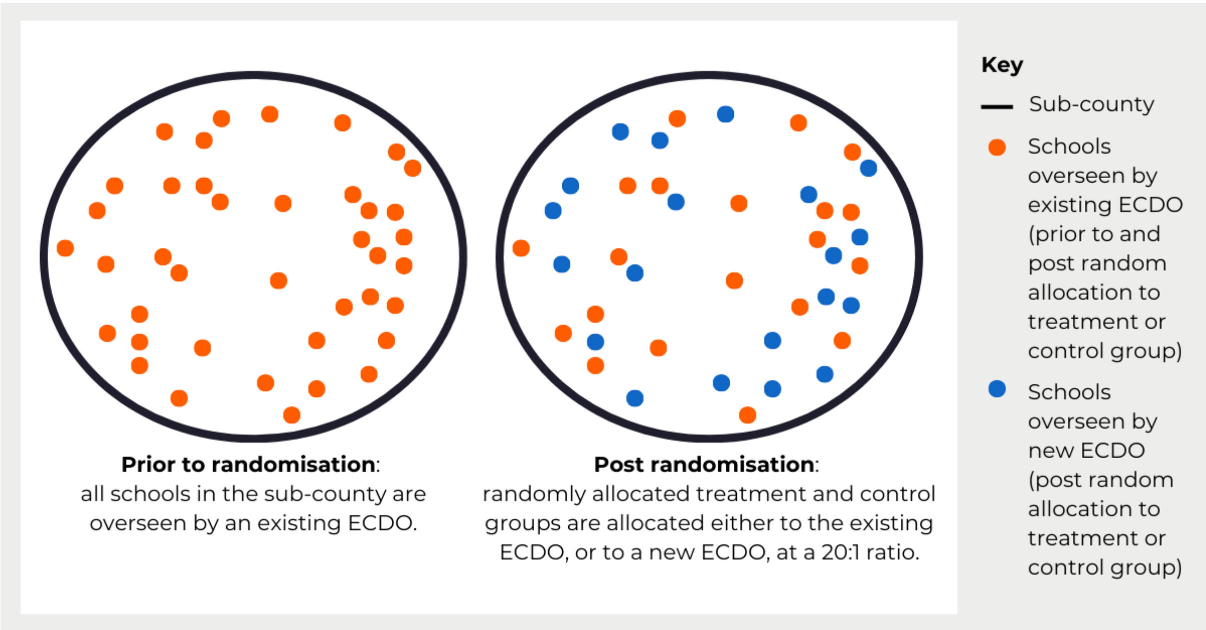
Recognising the complexity of RCTs in education research, it was anticipated that additional, randomly allocated schools may need to be added to the sample (e.g., due to the inability of a particular school to participate in the study for any ethical or contextual reason(s)). As such, the remaining 40 schools from the randomly sorted list of 316 were also randomly assigned to 'treatment' or 'control'. Schools were placed in a standby position according to the order of this randomly sorted list in case other schools had to be removed from the sample. Reasons for removal from a sample might be due to:

- an ethical reason identified prior to data collection commencing;
- the total number of learners turning out to be significantly lower than reported in school enrolment figures following further planning and investigation, and thus too low to satisfy the threshold of 10 assessments.

The process of potential randomised replacement took place in October 2022, prior to baseline assessment.

Once the full sample had been allocated to treatment and control groups, additional ECDOs were deployed within each sub-county to ensure a ratio of 20 schools to each ECDO (as agreed with the county government as part of the EIDU DPL-SP rollout strategy). Existing ECDOs (who had previously overseen all schools within their allocated sub-county) maintained relationships with all schools from either the treatment or control group in that sub-county. New ECDOs were assigned to the other group (i.e., either control or treatment). The geographic reach of the ECDO groups therefore remained unchanged, preventing any disruption to existing ECDOs. The only change was a reduction in the number of schools allocated to them. Additionally, the allocation of ECDOs prevented any contamination of treatment and control groups.

Figure 1. *Allocation of new ECDOs within the sub-county structure, to prevent contamination between the treatment and control group schools*



2.3.2 Learner-level randomisation (selection of learners for assessment)

Learners from PP1 classes in treatment and control groups were randomly selected for assessment at baseline. A list of random integers was generated for each school, using the same guidance observed for randomisation at the school level (↑J-PAL, no date). Two random lists of integers were generated for each school: one for girls, and one for boys.

On arriving at a school, enumerators were instructed to obtain attendance lists for all PP1 classes (likely alphabetised lists). In instances where multiple lists exist (due to there being more than one PP1 class), enumerators were instructed to randomly order the lists before applying the randomisation strategy across the entire PP1 grade.

Enumerators were then to select five girls for the assessment using the random integer list: i.e., if the first integer on this list is five, then they were to select the fifth girl on the attendance list, and so on. Up to 20 integers were provided on each list from which to select five girls, so the enumerator was able to continue selecting learners if several children were not in attendance. They were then to repeat this process, using the specific list of random integers for boys, to select five boys. As such, the randomised learner sample within each school cluster was stratified by gender.

In this way, 10 PP1 learners were randomly selected per school at baseline assessment. The same learners — if enrolled / in attendance — were then to be reassessed at midline and endline. Due to the timeline of the RCT straddling two academic years, it was anticipated that the majority of the learners would have transitioned to PP2 by midline and endline. To account for potential attrition over the course of the RCT, the sample size was increased by an extra 20% beyond that required to increase the likelihood that adequate statistical power would be achievable during final data analysis. In addition, a replacement strategy was designed for midline: if fewer than 10 learners were available for assessment at midline, then enumerators would use the same randomisation strategy (with the random integer list) to select additional learners until 10 learners were assessed. No replacement strategy will be employed for endline, with all learners from baseline and midline being reassessed as far as possible.

2.4 Learning assessment

The impact of the EIDU DPL-SP model on learning outcomes is being measured via three rounds of learning assessments, as outlined in Table 8.

A team of 42 enumerators was recruited and trained to conduct assessments using the IDELA assessment battery.

Table 8. *Timeline of RCT assessment rounds*

Assessment round	Dates	Time in relation to the Kenyan school year	Time in relation to treatment arm
Baseline	3–18 October 2022	The start of Term 3	Coinciding with the first three weeks of treatment rollout
Midline	15–26 May 2023	The start of Term 2	After two school terms of treatment arm (with one term of full DPL-SP model)
Endline	2–17 October 2023	The end of Term 3	After four school terms of treatment arm (with three terms of full DPL-SP model)

2.4.1 Assessment battery

IDELA is an assessment battery developed by Save the Children[6] which provides a holistic picture of children's development and learning covering four core developmental domains:

1. Motor development
2. Emergent literacy
3. Emergent numeracy
4. Social-emotional development.

IDELA is designed to assess children aged between 3–6 years in low-resource settings. The tool was validated via a series of analyses using data from 11 countries and 5,300 children (↑Save the Children, IDELA, 2023), demonstrating strong internal consistency (↑Pisani et al., 2018). Further exploratory and confirmatory factor analysis (↑Wolf et al., 2017) demonstrates that IDELA is a useful assessment tool for making inferences about children's school readiness. More recently, IDELA has been tested through studies in low- and middle-income country contexts in Africa and Latin America (↑Borzekowski et al., 2019; ↑Shavitt et al., 2022).

The assessment strategy for this RCT draws on the 'emergent numeracy' and 'emergent literacy' domains of IDELA. The assessment was estimated to take 14.5 minutes to complete (↑Seiden, 2021), meaning that a single enumerator can assess all 10 children from each school within a school day (a total of approximately 2.4 hours within a total school day of typically 3.5

[6] https://www.savethechildren.net/ Retrieved 8 October 2023

hours, from 8:30 am to 12:00 pm). However, if a child is unable to complete any of the first three tasks, then the assessment is ended early.

A full list of the assessment items available within the emergent numeracy and literacy domains is given in Table 9 below. The only adaptation made to the assessment battery was to replace the two 'letter names' items with equivalent items from the Measuring Early Learning Quality and Outcomes (MELQO) assessment tool.[7] This decision was taken because MELQO items are better aligned with the Kenyan CBC, with letter names and letter sounds being distinct in MELQO, unlike IDELA. As such, it responds to calls for RCTs to use outcome measures which align with the intervention and context, rather than broad outcome measures which may not produce useful evidence (↑Singer & Willett, 2003).

Table 9. *IDELA assessment items*

Domain	Item number	Item name
Emergent numeracy	1	Comparison by Size and Length
	2	Sorting and Classification
	3	Shape Identification
	4	Number Identification
	5	One-to-one Correspondence
	6	Addition and Subtraction
	7	Puzzle Completion
Emergent literacy	8	Expressive Vocabulary
	9	Print Awareness
	10	Letter Names (MELQO)
	11	Letter Sounds (MELQO)
	12	First Letter Sounds
	13	Emergent Writing
	14	Oral Comprehension

[7] https://www.ecdmeasure.org/what-is-melqo/ Retrieved 8 October 2023

The entire IDELA assessment battery was provided in both English and Kiswahili, with enumerators able to switch languages to aid children's comprehension. The battery was accessed by enumerators on the KoboCollect app via the KoboToolbox platform,[8] complemented by a pack of assessment stimulus cards and materials (e.g., printed picture cards, number / letter charts, buttons, and storybooks).

2.4.2 Enumerator recruitment, training, and schedule

An assessment team of 31 enumerators was recruited by WERK in September 2022, with a further 11 recruited over midline and endline to replace enumerators who were no longer available. Successful candidates were selected based on previous experience of:

A. Conducting learning assessments, such as the National Assessment System for Monitoring Learner Achievement (NASMLA)[9] or the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ).[10]

B. Qualitative and quantitative data collection, particularly with a focus on foundational literacy and numeracy in primary or pre-primary settings.

C. Experience with data collection tools, including ODK,[11] Kobo,[12] Tangerine,[13] or SurveyCTO.[14]

Ahead of baseline assessment, enumerators undertook four days of training, during which they received a detailed introduction to the focus of the research study, research ethics and safeguarding the assessment battery and baseline assessment logistics. At each subsequent assessment round, newly recruited enumerators received four days of training, while returning enumerators received two days of refresher training.

A core aspect of the training was the use of simulated inter-rater reliability (IRR) tests. Trainers conducted these tests by simulating a scripted assessment, and all enumerators scored the assessment at the same time. The IRR tests were then rapidly analysed (by the research manager and quantitative analysis advisor) and areas of variance were discussed in detail

[8] https://www.kobotoolbox.org/ Retrieved 18 October 2023
[9] https://www.knec.ac.ke/nasmla/ Retrieved 18 October 2023
[10] http://www.sacmeq.org/ Retrieved 18 October 2023
[11] https://getodk.org Retrieved 18 October 2023
[12] https://www.kobotoolbox.org/ Retrieved 18 October 2023
[13] https://www.rti.org/impact/tangerine-mobile-learning-assessments-made-easy Retrieved 18 October 2023
[14] https://www.surveycto.com/ Retrieved 18 October 2023

with the enumerators to improve the reliability of scoring across the team. While this did not feature the real-world variables of conducting IRR tests in the classroom, it had two benefits: first, it reduced the risk of the presence of multiple enumerators overwhelming the learners during assessments, thus potentially influencing the results of the assessment; and second, it enabled enumerators to receive rapid feedback on their approach prior to undertaking actual RCT assessments.

At each assessment round, enumerators were allocated a maximum of ten schools each, and they visited one school per day, over a period of two working weeks.

2.5 Supplementary questionnaires

To acknowledge that learning is influenced by various factors and does not occur in isolation, supplementary questionnaires were created to collect data on learner characteristics, teachers, and the school environment. This was done for both treatment and control groups. Additional questions were also designed as proxies for possible socio-economic factors affecting learning outcomes. Much of the data was provided by the teachers or from the attendance register, with several short questions tailored specifically to the learner. The presence or absence of key EIDU resources in the classroom was also assessed in treatment schools. A summary of supplementary data collected is given in Table 10.

Table 10. *Overview of supplementary characteristics data*

Domain	Characteristics data gathered
Learners (source: attendance register, teacher identification, enumerator observation, learner self-reporting)	■ Age ■ Gender ■ Any physical or learning disability ■ Whether the learner has a school uniform ■ Whether the learner is wearing shoes ■ Learner's teacher ■ Language spoken at home ■ Whether home leaks when it rains ■ Water source at home ■ Availability of a mobile device at home
Teachers (source: school records, teacher self-reporting, enumerator observation)	■ Gender ■ Single- or multi-grade teacher ■ Years of teaching experience ■ Highest level of education ■ Highest level of teacher training ■ Access to: Learner progress record; Scheme of work; Record of work; Lesson plan; Health records; Syllabus/curriculum design

	■ Size of class ■ Language of instruction ■ Length of commute (minutes)
Schools (source: school records, teacher identification)	■ Standalone or part of a primary school ■ Presence of PP1 and/or PP2 classes ■ Other interventions operational at the school ■ The provision of school meals ■ The availability of running water ■ The availability of electricity ■ Geographic features in immediate vicinity of the school ■ Provision of refuse collection ■ Distance to nearest market ■ Distance to nearest health clinic or hospital ■ Most common occupations of parents
Treatment school data (source: enumerator observation)	■ The availability of a functional or non-functional EIDU mobile device ■ The availability of a charging cable for the mobile device ■ The presence of a labelled EIDU corner

The questionnaires were administered by enumerators on the same day as the assessments were undertaken at each school. Data was collected via KoboCollect.

In addition, three Likert scale questions for teachers in the treatment groups were designed for endline assessment. These questions were perception-based, to provide an initial indication of whether teachers' personal experience and views of the EIDU DPL-SPL model correlated with its impact on learning outcomes. The three questions covered the extent to which teachers feel:

1. the Tayari lesson plans (on the EIDU app) are helpful for their teaching;
2. the digital activities on the EIDU app (i.e. the DPL tool) are helpful for students' learning;
3. they are satisfied with the EIDU app.

2.6 Integration with EIDU app data

While the primary dataset for the RCT is the three rounds of learning assessments using the IDELA battery, the treatment arm also offers the opportunity to test whether IDELA data correlates with the continuous learning data collected via the EIDU app.

Interspersed among learning exercises on EIDU's DPL tool are a set of five assessment exercises, which are digitised numeracy and literacy assessment items from assessment batteries, including early grade reading assessment (EGRA),[15] early grade mathematics assessment (EGMA),[16] and MELQO. Every six weeks, the EIDU app integrates these assessments alongside other learning content for each individual learner. The assessments generate detailed learning outcomes data, including the number of correct / incorrect inputs by the learner during each individual exercise, as compared to the general DPL exercises, which produce an overall score as a summative evaluation of learning. In addition, formative assessment data is available from learners' engagement with the other learning exercise on the DPL tool, which can also be analysed to investigate the relationship between in-app performance and IDELA assessment outcomes.

This data is collected anonymously on the EIDU app (via unique user IDs associated with each learner). By associating these IDs with the anonymous learner codes generated for each RCT assessment, it is possible to link each individual IDELA score with that learner's learning history on the EIDU app. The data that is shared between EIDU and the research team is fully anonymised to avoid the sharing of personal data and to ensure learners in the sample cannot be identified. The analytical strategy for comparing these two datasets is outlined below, along with the analytical strategy for the main RCT.

2.7 Analysis strategy

The primary focus of the analytical strategy will be to attend to the longitudinal nature of the RCT: to examine how assessment scores changed over the course of the three assessment rounds and the difference in learning outcomes between the treatment and control schools. In addition, further analysis will consider factors which may be influencing learning outcomes and the impact of the EIDU DPL-SP model.

2.7.1 Data cleaning

The full data set will undergo a thorough cleaning process (coordinated by the research manager, with input from the research associates and quantitative analysis advisor). Cleaning will be systematically documented, following three key steps:

[15] https://www.edu-links.org/resources/early-grade-reading-assessment-egra-toolkit Retrieved 18 October 2023
[16] https://www.edu-links.org/resources/early-grade-mathematics-assessment-egma-toolkit Retrieved 18 October 2023

1. Entries will be removed from the dataset for ethical reasons. For example, if a child has withdrawn their consent during the assessment or become overwhelmed and hence ended the assessment early.

2. All entries will be cleaned (removing duplicates; recoding blanks as NA; recoding 'other' entries as new variables, where relevant; etc.) and carefully cross-referenced (e.g., resolving inconsistency errors, such as between learner name and learner code). Any entries where issues with the data cannot be resolved will be removed.

3. Teacher and learner data will be merged, so that the assessment data can be disaggregated by variables, including school and teacher characteristics. Data on these characteristics were collected via a different survey tool, to avoid duplication with every learner assessment, but can be linked to individual assessment via a unique code generated for each teacher, and input into both survey tools.

2.7.2 Calculating aggregate IDELA scores

The IDELA assessment battery is structured into four levels: the overall IDELA score comprises two domains (numeracy and literacy); each of these domains comprises seven items (see Table 9 above); and each of the items is made up of sub-tasks. This method is in line with previous research on IDELA where the percentage of correctness is calculated per item and then averaged to domain and overall scores unweighted (↑Pisani et al., 2018).

To create aggregate assessment scores, the following steps will be taken:

- At the sub-task level, correct tasks will be coded as 1, incorrect tasks as 0 and any skipped tasks will be coded as N/A to reflect the missing data.

- At the item level, percentage correctness will be calculated across sub-tasks, treating N/A as 0 if the learner answers at least one sub-task within that item. This analytical approach is recommended by Save the Children.

- The same approach will be adopted to calculate the domain score for numeracy and literacy: 'N/A' item scores (those where every sub-task is N/A) will be treated as 0 in order to generate an aggregate domain score.

- Finally, to obtain the overall IDELA score, the domain scores will be averaged by treating all N/As as 0.

Calculation of IDELA scores will take two further considerations into account. First, the persistence and engagement subtasks (in Items 5, 7, and 14) will be analysed separately to the other subtasks in those items. This is due to the subjective nature of these 'self-regulation' assessment metrics and is in accordance with the IDELA Adaptation & Administration Guide, which stipulates that these subtasks are optional and not part of the core IDELA tool. As such, the scores for these subtasks will not contribute to the percentage correctness for the overall item.

Second, analysis of the IDELA scores will consider the possible weighting of item scores, based on low-response rates and/or the identification of specific sub-tasks, which may contribute to skews in the aggregate IDELA scores. This will build directly on the findings of other IDELA evaluations. For instance, studies by ⇡Rey-Guerra et al. (2022) and ⇡Wolf et al. (2017) removed low-response subtasks, such as the second set of 10 subtasks in Item 4. ⇡Wolf et al. (2016) also indicate the loadings of the different IDELA subtasks, including the high loadings of letter and number identification items. These and other findings will be taken into account in the analysis of IDELA data for this RCT.

2.7.3 Analysis of cross-sectional group difference

Group comparison will examine the difference between the control and treatment groups to reveal the effects of the EIDU DPL-SP model on learning outcomes. Independent sample t-tests (⇡Lohr et al., 2014; ⇡Ross et al., 2017; ⇡Ruble et al., 2012) will be used to test the group difference. The analysis of group difference will also incorporate effect size calculation (see Section 2.7.5). This approach has been used in other IDELA evaluations, including that of ⇡Shavitt et al. (2022) who used t-tests and ANOVA (as well as corresponding non-parametric analysis) to check group differences based on gender, age (4 years vs 5 years), and maternal education level.

Building on the exploratory results from independent t-tests, multilevel modelling will be used to add in other variables and account for any between-school and within-school differences. This is a suitable approach for a cluster RCT, enabling it to account for any dependency between learners' performance and factors associated with their school cluster (⇡Niehaus et al., 2014; ⇡Raudenbush & Bryk, 2002). Multilevel modelling has been adopted by previous RCT research on early grade learning outcomes (e.g. ⇡Bang et al., 2023; ⇡Thai et al., 2022) and in research using the IDELA assessment battery (e.g., ⇡Maldonado-Carreño et al., 2022).

2.7.4 Longitudinal analysis across time points

One important factor of the longitudinal nature of the study is the retention of learners between and across time points (see previous research on Tayari, ↑Piper et al., 2018b). Paired sample t-tests will be used to analyse students' learning gains between assessment points for learners who were at baseline and midline, midline and endline, as well as baseline and endline.

As with the analysis of cross-sectional group difference, multilevel modelling can be used to conduct longitudinal analysis. This will analyse learning progression across three time points (↑Dedrick et al., 2009; ↑Peugh, 2010), assuming there is a sufficient sample size of learners who participated in all three assessments. Other methods — including random slopes and mixed-effects models — will also be considered.

Analysis will also consider whether the treatment arm experienced ceiling effects, as indicated in a study by (↑Piper et al., 2018a). If ceiling effects are observed (↑Garin & Michalos, 2014), this may suggest that there is limited scope for learners to improve their scores. The focus will be on learners who have already attained high scores at midline and whether they continue to further improve their scores at endline. It is anticipated that the uplift of scores between midline and endline will be smaller compared to the uplift from baseline to midline.

2.7.5 Calculation of effect size

Effect sizes will be calculated to assess the impact of the EIDU DPL-SP model on learning outcomes, providing a standardised magnitude of difference between the treatment and control groups. Effect sizes will be calculated for each of the 14 items in the numeracy and literacy domains, as well as for each aggregated domain and overall IDELA score. Effect sizes will be further stratified by gender, sub-county and other demographic variables.

Cohen's *d* will be the primary approach to calculate effect size, employing pooled standard deviation. This is an approach adopted in similar cluster RCT studies (↑Ngware et al., 2018; ↑Yousafzai et al., 2018) and has been frequently discussed in the literature (e.g., ↑Durlak, 2009 for the formulas). Although conventional interpretations of Cohen's *d* suggest that 0.2 is a 'small' effect size, it has been argued that, in educational research, an effect size of 0.2 could have policy implications (↑Hedges & Hedberg, 2007). In this study, the interpretation of Cohen's *d* will closely reference relevant results from existing studies in similar contexts (e.g., ↑Ngware et al., 2018,

↑Pitchford, 2015), while using the conventional benchmarks of 0.2 (small), 0.5 (medium), 0.8 (large) as a reference.

There is a lack of consensus in the literature on how to calculate effect size in multilevel modelling. Two approaches are offered: global effect sizes can indicate the variance of the outcome variable, using all the predictors in the model; local effect sizes, on the other hand, indicate variance based on the level 1 (i.e., the lowest level) predictors (↑Peugh, 2010). ↑Singer & Willett (2003) recommend pseudo-R^2 as the index of global effect size. Local effect size calculation is exemplified by ↑Bryk & Raudenbush (1992), by estimating the proportion of variance reduction after adding in level 1 predictors. While Cohen's *d* will provide a standardised magnitude of effect for this study, these additional approaches to calculating effect size for multilevel modelling will also be explored.

2.7.6 Attrition analysis

Attrition analysis is recommended in RCT studies using IDELA in low- and middle-income countries (↑Ngware et al., 2018; ↑Piper et al., 2018b). Such analyses of the data on the assessment and characteristics of learners who were only available at baseline assessment but dropped out before midline or endline will be undertaken. This will help to indicate whether any particular factors influenced attrition and skewed the assessment data across the RCT.

2.7.7 Prediction of learning outcomes

Studies have explored whether certain characteristics of students, teachers, and schools play a role in learning (↑Piper et al., 2018b; ↑Shavitt et al., 2022). Assessment scores will be regressed by learner, teacher, and school demographic data. This will indicate which variables predict learning across time and, therefore, which subgroup of learners is most likely to benefit from the EIDU DPL-SP model. In other words, whether there is a disproportionate impact for certain subgroups of learners, relative to their comparable subgroup in the control group. For instance, the analysis conducted by ↑Shavitt et al. (2022) indicated possible influencing factors such as gender, age, and maternal education level. In another IDELA study, ↑Maldonado-Carreño et al. (2022) investigated factors such as teacher education, teaching experiences, and school environment. In this vein, the analysis strategy involves investigating which, if any, of the demographic or perceived factors (e.g., gender, language of instruction, access to electricity (see the full list in Table 10 above)) potentially influence learning. Stratification of learning outcomes by classroom size, as well as an exploratory weighting of scores by this variable, will further investigate whether the class size is an influencing factor, and to account for the fact that the number of assessments per school was not proportional to grade / class size.

2.7.8 Analysis of the EIDU app data

Through the merging of the IDELA assessment data with the learning data generated on the EIDU app, it will be possible to test whether the two data sets corroborate each other in learning outcomes. Correlation analysis can be conducted to investigate whether the IDELA assessment scores are consistent with data on the EIDU app (which includes assessment scores and non-assessment measures, including usage time).

2.7.9 Cost-effectiveness analysis

An initial estimate of cost-effectiveness will also be achieved by linking effect size calculations to projected costs per learner. The analysis will involve a calculation of (micro-) Learning-Adjust Years of Schooling (LAYS), drawing on the work of ↑Angrist et al. (2020), ↑Evans & Yuan (2019), and ↑Islam et al. (2022). First, equivalent years of schooling, *e*, will be determined using the learning gains for the control and treatment groups, expressed as standard deviation gains:

$$e = \frac{\beta_i^{\sigma, T}}{\delta_{i,X}^{\sigma, T}}$$

Beta is learning gain in the treatment group in terms of standard deviation; delta is learning gain in the control group in terms of standard deviations. Since the comparison will be drawn against a control group (rather than the population in general), this metric will generate micro-LAYS rather than LAYS (see ↑Evans & Yuan, 2019).

Second, the learning adjustment factor, L_i^h, will be calculated by dividing the harmonised test score for Kenya by a benchmark (e.g., Singapore):

$$L_i^h = \frac{\delta_i}{\delta_h}$$

This is not dependent on the RCT data and stands at 0.79.

Duration adjustment, *t*, will take into account that the RCT intervention is taking place over four terms — longer than a standard Kenyan school year.

Finally, LAYS will be calculated by multiplying the three components:

$$LAYS^{l} = e * L_{i}^{h} * t$$

LAYS per USD 100 will then be calculated using this and the cost data. Further approaches to calculating cost-effectiveness, including factoring in aspects such as equity (↑Sabates et al., 2021), may also be considered.

3. Methodological limitations

While the research design aims to consider contextual realities, it is recognised that there are limitations associated with the constantly changing context of education. Furthermore, it is necessary to acknowledge that the field of education, ranging from national systems to the classroom level, differs from a clinical setting. This implies there are methodological limitations associated with conducting an RCT in this context. These limitations, and mitigating steps taken, are outlined below.

3.1 Contextual relevance of RCT conditions

As outlined by ↑Sullivan (2011), there is a risk that in seeking to achieve randomisation, education RCTs may sacrifice the relevance of the research findings to the broader sector. This is due to the fact that, although randomisation may reduce allocation bias derived from the underlying characteristics of the sample population, it cannot control other sources of error which may occur in education research. Reported best practice regarding successful scaling of educational interventions (e.g., ↑Piper et al., 2018a) has provided a foundation from which the RCT has been designed and will be considered when disseminating findings.

3.2 School-level randomisation

A key challenge was randomisation at the school level. This was due to the deployment and training of ECDOs at the sub-county, which risks contamination of the control group (i.e., an ECDO implementing the treatment approach in control schools). It was not possible to randomise at the sub-county level (within the existing allocation of ECDO groups), as this demanded a very large sample size to achieve appropriate randomisation (beyond the capacity of presently allocated resources).

The selected approach to randomisation (assigning new ECDOs to schools within each sub-county at a ratio of 20:1) attempted to mitigate any contamination between groups. Crucially, in July 2022, representatives from the Ministry of Education, Council of Governors (COG), and KICD, along with representatives from 46 out of 47 counties in Kenya officially adopted resolutions to adopt an ECDO-to-school ratio of 20:1 in all 46 counties to which the DPL-SP model will be rolled out. Therefore, the relevance of this chosen randomisation approach to scaling the intervention across the rest of the country is high. However, it is recognised that potential limitations remain.

Contamination The issue of contamination is complicated by the geographical proximity of some schools (which may have been randomly assigned to different groups) to each other. This required a clear communication strategy to mitigate confusion among the ECDOs: the reasoning for the new group assignments was clearly articulated during ECDO training.

Community tension Due to the nature of the randomisation strategy for the RCT, some schools inside existing ECDO groups were selected for treatment while others were not. This required a similar mitigation strategy to the limitation mentioned above. A clear process of communication to all stakeholders involved (including school teachers, headteachers, ECDOs, and county government officials) about the phased nature of the DPL-SP rollout and the reason some schools were selected for the treatment groups before others were necessary. It also required a clear ethical justification for the choice of an RCT and a proposed future plan for control schools — namely, that all schools in these sub-counties will receive the treatment via a phased rollout starting in January 2024 (see Section 4).

ECDO selection Whether a new or existing ECDO received the Tayari training was considered to be another source of possible tension between ECDOs in the same sub-county. Again, the justification for this and the timeline for all officers / schools to receive Tayari training was communicated during ECDO training. Additionally, it was unknown whether the learning impact would be higher from existing ECDOs getting Tayari training or whether new ECDOs would be more effective. To mitigate this, new and existing ECDOs were split between the treatment and control groups as evenly as possible. A balance of 12 ECDOs supporting the treatment group (six existing and six new officers) and 8 ECDOs supporting the control group (four existing, who act as programme officers for the whole sub-county, and four new officers) was achieved.

For practical reasons (i.e., to ensure a sufficient number of participants), those schools with fewer than 15 learners in either PP1 or PP2 classes were excluded from the randomisation process. This means schools considered to be very small were excluded from the analysis. Further, the appropriateness of clustering schools geographically within sub-counties, and then using clusters of treatment and control groups, was explored but decided against in order to prevent potential skewing by geographical location.

3.3 Learner-level randomisation

Learners transitioning from PP1 to PP2 during the timeline of the RCT were selected to be the participants of the RCT. Randomisation to select the ten learners for assessment at baseline therefore took place at the PP1 grade level. However, this was complicated by the fact that some schools only have one class at PP1, while others have multiple classes.

It was therefore decided that, in the case of multiple PP1 classes within a single school, randomisation would be achieved by applying the list of random integers to a combined attendance list from all classes. The alternative of selecting a single class in which to do randomisation would have simplified the process of randomisation. However, this was nonetheless deemed to be more of a limitation since there was no guarantee that learners would remain in the same class at PP2. Furthermore, learners may be assigned classes via 'streaming' (assigning classes based on learning level), which would result in a skewed sample.

It is recognised, however, that there were potential limitations to achieving learner-level randomisation across a whole grade, rather than within an individual class. Although as outlined in Section 2.5, teacher names and demographic information were collected to enable potential subgroup analyses to explore impacts on learning.

Finally, having 10 children per school meant that all calculations assume that schools are the same size. It is recognised that school size and classroom size may be an important factor in the effectiveness of the EIDU DPL-SP model, and as such, stratification and possible exploratory weighting has been built into the analytical strategy to account for class size.

3.4 Phased rollout of treatment arm

EIDU generally leave a 6–12 month gap between schools receiving the DPL functionality and the digital Tayari content. This allows teachers and learners to familiarise themselves with the first part of the intervention before they receive the second.

However, in the timeline of this RCT, the gap between the two parts of the intervention was condensed to 3 months. This was done to accommodate the time constraints of the RCT (enabling teachers / learners to use both functionalities for 9 months before endline assessment). While this slightly changed usual practice, it was not considered to be a limitation, since teachers and learners had a whole school term to use the DPL functionality before being given the Tayari content. This was deemed to be long enough to gain familiarity.

3.5 Participant engagement in the research

Ensuring participant engagement in the research (including teachers' and headteachers' support of assessment at three proposed data points over a year's period) is crucial to enable the proposed sample to be achieved and reduce the risk of participant attrition.

To reduce this risk, headteachers were contacted by enumerators in advance of their assessment visits to inform them of the date, time, and purpose of their visit. Furthermore, on arrival, enumerators would present the headteacher with research documents. These included the research permissions from the National Commission for Science, Technology and Innovation (NACOSTI), Kenya Medical Research Institute (KEMRI), and the county government) and an information sheet about the research. This underpinned the perceived legitimacy of the research and facilitated greater ease of communication and relationship with the stakeholders at the sample schools.

3.7 Logistical challenges of assessment

The assessment data was recorded on KoboCollect — an app downloaded onto digital tablets. The associated risks of this approach were twofold: that the data would not be correctly uploaded and that the devices would run out of battery, disrupting assessment. The first of these risks was mitigated by utilising the offline version of Kobo, which enables users to input data even when they have no internet connection, and then upload the data at a later date. The second was mitigated by the fact that approximately one hour of flexible time was built into the assessment day, enabling enumerators to charge their devices for a short period if required. Enumerators were also provided with a backup paper tool and data entry template, in case the tablet malfunctioned.

A second challenge was the issue of language of instruction. While the IDELA assessment tool was provided in English and Kiswahili (allowing enumerators to switch codes while communicating with learners), teachers in Murang'a also teach in the language of the school's catchment area (including Kikuyu, Kamba, and Gusii). While this is a potential limitation of the assessment approach, it was anticipated that the vast majority of teachers also teach in Kiswahili (a slightly lower proportion in English), ensuring that learners should have some familiarity with literacy and numeracy content in both languages. However, in order to assess the impact of language(s) of instruction and language spoken by learners at home, additional questions were asked via the supplementary questionnaires to associate these variables with the assessment data (see Table 10), and are part of the analytical strategy (see Section 2.7.7).

4. Ethical considerations

RCTs in social settings, including the education sector, pose fundamental ethical dilemmas, in particular, concerning the methodological need for a control group. The steps taken to address these ethical considerations are outlined in this section.

Research ethics also encompass the four basic principles of the participants' right to informed consent, anonymity, confidentiality, and privacy. All members of the research team are aware of EdTech Hub's ethical approval, safeguarding, and risk management policies with appropriate training being undertaken where appropriate / required (e.g., on human and social research ethics). Specific considerations related to each of these principles are further outlined below.

4.1 The ethics of randomised controlled trials

An ethical concern in the use of RCTs in education research is the need for a control group, who will be deprived of the potential benefits of the intervention. Following an RCT it may be desirable to design a short catch-up intervention for the control group participants should effectiveness be demonstrated. While the ethical implications of using control groups are recognised — i.e., that the schools in the control group did not receive EIDU's DPL-SP model in 2022/23 — the phased nature of the EIDU rollout somewhat tempers this concern. Due to the operational impracticalities of rolling out the tool to all schools at once, EIDU have staggered the rollout to pre-primary schools in Embu, Murang'a, and Kiambu counties across 2022 and 2023. Therefore, control schools in these counties are not the only schools excluded from receiving the tool simultaneously with the treatment schools participating in the RCT. As part of the staggered rollout, they will receive the tool from the start of 2024.

4.2 Informed consent

Gatekeeper consent has been sought for children participating in the study: teachers of learners involved in the RCT assessment and mixed-methods research were asked to complete a gatekeeper consent form at each assessment round, providing permission on behalf of parents and caregivers. Parents and caregivers were informed of the research through information sheets distributed by the headteacher. In addition, at the school level, verbal assent was also requested from each pupil before they were assessed. Assent is a verbal agreement to participate in research freely and without coercion given by an individual who is under the legal age and implies that they have agreed to participate in the study without undue influence or pressure. The assent statement was included as an introductory step for each learner assessment, with the assessor reading the statement to each child in English or Kiswahili.

Written consent was additionally sought from the teachers involved in the RCT. A consent form was provided by the researchers prior to the start of the teacher questionnaire and consent was recorded on the questionnaire form, as well as on a signed consent form.

There was no provision for benefits directly for individual children, other than the receipt of a reward of a pencil and eraser.

4.3 Safeguarding

A comprehensive coverage of research ethics and safeguarding policies was included as part of the four-day training for new enumerators and a two-day refresher training for returning enumerators at each assessment round.

In recognition that the RCT involves particularly vulnerable participants (learners aged 4–6), the logistics of assessment were designed to prioritise the safety of participants. Enumerators were instructed to conduct assessment in a quiet location which minimised distractions (such as having other adults present who may interfere with the assessment), but which was nonetheless within sight of another adult. This was specified to ensure that no child was left alone with an enumerator.

Furthermore, it is recognised that the possibility of psychological or emotional harm exists when collecting data among such young children. To mitigate against such harms, research participants were required to voluntarily participate in the study — learners were able to withdraw if they were no longer interested or able to participate.

4.4 Anonymity, confidentiality, and privacy

All research partners (EdTech Hub, WERK, and EIDU) will protect the anonymity of all participants in the proposed research. This involves the non-disclosure of the identity of participants, including their names, description of physical appearance or association with any institution. Each child has been assigned a unique identifier or code during assessment, which will be used during the analysis and presentation of data to protect anonymity.

Confidentiality refers to the handling of identifiable data about a research participant and to agreements about how such information is to be stored in keeping with the participants' interests, especially in controlling the access of others to such data. In the present research, confidentiality of information is guaranteed by storing all hard copies of completed instruments in a non-accessible locked area and archiving password-protected soft copies of all data keyed into computers. The data is only accessible to the research team.

Privacy refers to the participants and their interest in controlling access to them by other people. Privacy is an extension of the principle of confidentiality and refers to the setting, style, and circumstances under which an investigator obtains information from participants. Researchers are aware of the various research procedures and methods that can be used to respect privacy, where other people are not likely to be watching or listening to very personal information. All research partners are also attending to the privacy of participants. Raw data is only accessible to the immediate research team.

The findings from the research will be analysed and shared, both within the research community and specifically with the county and national government from a policy perspective to improve the quality of instruction. This ultimately intends to benefit the learner and others like them. All findings will be anonymised.

4.5 Data management

All data related to the research are stored securely in a password-protected electronic database, with access limited to those directly involved in the research. All hard copies of data are stored securely or will be destroyed once copies are uploaded to the electronic database. All data will be anonymised during the presentation of findings, using random unique identifiers to remove all inferences of participants' identities.

4.6 Ethics Review Committee

An application for ethical approval for the first phase of the research (design-based research in Mombasa) was led by EIDU and granted by the Scientific and Ethics Review Unit (SERU), a unit in the Kenya Medical and Research Institute (KEMRI - REF: NON-KEMRI PROTOCOL NO. 4444), NACOSTI (Licence No: NACOSTI/P/22/17399) and the Ministry of Education. An extension of this research approval was then granted for the second phase of the research (the RCT in the four sub-counties of Murang'a, as outlined in this document) from the same parties. An application was also approved by the EdTech Hub ethics panel.

EIDU is working in collaboration with county government officials in advance of the rollout and informing them of any changes to the research plan before any future stages of the research.

References

These references are available digitally in our evidence library at https://docs.edtechhub.org/lib/JNH4277Z

Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, F. H., & Sabarwal, S. (2020). *How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions using the New Learning-Adjusted Years of Schooling Metric*. The World Bank. https://doi.org/10.1596/1813-9450-9450. Available from https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9450. (details)

Banerjee, A., Andrab, T., Banerji, R., Dynarski, S., Glennerster, R., Grantham-Mcgregor, S., Muralidharan, K., Piper, B., Saavedra Chanduvi, J., Yoshikawa, H., Ruto, S., & Schmelkes, S. (2023). *2023 Cost-effective Approaches to Improve Global Learning — What does Recent Evidence Tell Us are "Smart Buys" for Improving Learning in Low- and Middle-income Countries?* World Bank Group. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099420106132331608/IDU0977f73d7022b1047770980c0c5a14598eef8. (details)

Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an Adaptive Game-Based Math Learning App to Support Personalized Learning and Improve Early Elementary School Students' Learning. *Early Childhood Education Journal*, *51*(4), 717–732. https://doi.org/10.1007/s10643-022-01332-3. (details)

Borzekowski, D. L. G., Lando, A. L., Olsen, S. H., & Giffen, L. (2019). The impact of an educational media intervention to support children's early learning in Rwanda. *International Journal of Early Childhood*, *51*(1), 109–126. https://doi.org/10.1007/s13158-019-00237-4. (details)

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc. (details)

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016. *Educational Research*, *60*(3), 276–291. https://doi.org/10.1080/00131881.2018.1493353. (details)

Coskinas, X., Simes, J., Schou, M., & Martin, A. J. (2020). Changes to aspects of ongoing randomised controlled trials with fixed designs. *Trials*, *21*(1), 1–8. https://doi.org/10.1186/s13063-020-04374-3. (details)

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research*, *79*(1), 69–102. https://www.jstor.org/stable/40071161. (details)

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928. https://doi.org/10.1093/jpepsy/jsp004. (details)

Engeström, Y. (2011). From design experiments to formative interventions. *Theory & Psychology*, *21*(5), 598–628. https://doi.org/10.1177/0959354311419252. (details)

Evans, D. K., & Yuan, F. (2019). *Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms*. https://doi.org/10.1596/1813-9450-8752. Available from http://hdl.handle.net/10986/31315. (details)

Friedberg, A. (2023). Can A/B Testing at Scale Accelerate Learning Outcomes in Low- and Middle-Income Environments? In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (pp. 780–787). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_119. (details)

Garin, O. (2014). Ceiling Effect. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 631–633). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_296. (details)

Global Education Monitoring Report Team. (n.d.). *Global education monitoring report, 2023: technology in education: a tool on whose terms?* Retrieved October 12, 2023, from https://doi.org/10.54676/UZQV8501. (details)

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706. (details)

Islam, A., Wang, L. C., & Hassan, H. (2022). *Delivering Remote Learning Using a Low-Tech Solution: Evidence from an RCT during the Covid-19*

Pandemic. EdTech Hub. https://doi.org/10.53832/edtechhub.0070. Available from https://docs.edtechhub.org/lib/FE3VBQQW. (details)

J-PAL. (n.d.). *Exercise B: How to do random assignment using MS Excel*. Abdul Latif Jameel Poverty Action Lab. https://www.povertyactionlab.org/sites/default/files/research-resources/ExerciseB_RandomizationMechanics_Excel_Instructions_Final.pdf. (details)

Kendall, J. M. (2003). Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, *20*(2), 164–168. https://doi.org/10.1136/emj.20.2.164. Available from https://emj.bmj.com/content/20/2/164. (details)

Kenya National Bureau of Statistics. (2019). 2019 Kenya Population and Housing Census Volume IV: Distribution of Population by Socio-Economic Characteristics. *Kenya National Bureau of Statistics*. https://www.knbs.or.ke/download/2019-kenya-population-and-housing-census-volume-iv-distribution-of-population-by-socio-economic-characteristics/. (details)

Lohr, S., Schochet, P. Z., & Sanders, E. (2014). *Partially Nested Randomized Controlled Trials in Education Research: A Guide to Design and Analysis*. National Center for Education Research (NCER). https://files.eric.ed.gov/fulltext/ED545532.pdf. (details)

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850. (details)

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, *52*(5), 1935–1964. https://doi.org/10.1111/bjet.13116. Available from https://onlinelibrary.wiley.com/doi/10.1111/bjet.13116. (details)

Maldonado-Carreño, C., Yoshikawa, H., Escallón, E., Ponguta, L. A., Nieto, A. M., Kagan, S. L., Rey-Guerra, C., Cristancho, J. C., Mateus, A., Caro, L. A., Aragon, C. A., Rodríguez, A. M., & Motta, A. (2022). Measuring the quality of early childhood education: Associations with children's development from a national study with the IMCEIC tool in Colombia. *Child Development*, *93*, 254–268. https://doi.org/10.1111/cdev.13665.

Available from https://srcd.onlinelibrary.wiley.com/doi/10.1111/cdev.13665. (details)

McNulty, K. (2021). *Handbook of Regression Modeling in People Analytics: With Examples in R and Python*. CRC Press. (details)

Ngware, M., Hungi, N., Wekulo, P., Mutisya, M., Njagi, J., Muhia, N., Wambiya, E., Donfouet, H., Gathoni, G., & Mambe, S. (2018). *Impact evaluation of Tayari school readiness program in Kenya*. APHRC. https://aphrc.org/wp-content/uploads/2019/07/Impact_Evaluation_ECDE_Tayari-long-report.pdf. (details)

Niehaus, E., Campbell, C. M., & Inkelas, K. K. (2014). HLM behind the curtain: Unveiling decisions behind the use and interpretation of HLM in higher education research. *Research in Higher Education*, *55*(1), 101–122. https://doi.org/10.1007/s11162-013-9306-7. (details)

Orkin, A. M., Gill, P. J., Ghersi, D., Campbell, L., Sugarman, J., Emsley, R., Steg, P. G., Weijer, C., Simes, J., Rombey, T., Williams, H. C., Wittes, J., Moher, D., Richards, D. P., Kasamon, Y., Getz, K., Hopewell, S., Dickersin, K., Wu, T., … CONSERVE Group. (2021). Guidelines for reporting trial protocols and completed trials modified due to the COVID-19 pandemic and other extenuating circumstances: The CONSERVE 2021 Statement. *JAMA*, *326*(3), 257–265. https://doi.org/10.1001/jama.2021.9941. (details)

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*(1), 85–112. https://doi.org/10.1016/j.jsp.2009.09.002. (details)

Piper, B., Destefano, J., Kinyanjui, E. M., & Ong'ele, S. (2018a). Scaling up successfully: Lessons from Kenya's Tusome national literacy program. *Journal of Educational Change*, *19*(3), 293–321. https://doi.org/10.1007/s10833-018-9325-4. (details)

Piper, B., Sitabkhan, Y., & Nderu, E. (2018b). Mathematics from the beginning: Evaluating the Tayari preprimary program's impact on early mathematics skills. *Global Education Review*, *5*(3), 57–81. https://ger.mercy.edu/index.php/ger/article/view/434. (details)

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research*, *91*, 1–15. https://doi.org/10.1016/j.ijer.2018.06.007. Available from https://www.sciencedirect.com/science/article/pii/S0883035518301885. (details)

Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00485. (details)

Pitchford, N. J., Chigeda, A., & Hubber, P. J. (2019). Interactive apps prevent gender discrepancies in early-grade mathematics in a low-income country in sub-Sahara Africa. *Developmental Science*, *22*(5), 1–14. https://doi.org/10.1111/desc.12864. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.12864. (details)

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed). Sage Publications. (details)

Republic of Kenya Ministry of Education. (2017). *National Pre-Primary Education Policy 2017*. https://repository.kippra.or.ke/handle/123456789/560. (details)

Republic of Kenya Ministry of Education. (2018). *National Pre-primary Education Policy Standard Guidelines*. https://planipolis.iiep.unesco.org/sites/default/files/ressources/pre-primary_policy_guidelines_11_1.pdf. (details)

Rey-Guerra, C., Maldonado-Carreño, C., Ponguta, L. A., Nieto, A. M., & Yoshikawa, H. (2022). Family engagement in early learning opportunities at home and in early childhood education centers in Colombia. *Early Childhood Research Quarterly*, *58*, 35–46. https://doi.org/10.1016/j.ecresq.2021.08.002. Available from https://www.sciencedirect.com/science/article/pii/S0885200621000946. (details)

Ross, A., & Willson, V. L. (2017). Independent Samples T-Test. In A. Ross & V. L. Willson (Eds.), *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures* (pp. 13–16). SensePublishers. https://doi.org/10.1007/978-94-6351-086-8_3. (details)

Ruble, L., McGrew, J. H., & Toland, M. D. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of Autism and Developmental Disorders*, *42*(9), 1974–1983. https://doi.org/10.1007/s10803-012-1446-7. (details)

Sabates, R., Rose, P., Alcott, B., & Delprato, M. (2021). Assessing cost-effectiveness with equity of a programme targeting marginalised girls in secondary schools in Tanzania. *Journal of Development*

Effectiveness, *13*(1), 28–46. https://doi.org/10.1080/19439342.2020.1844782. (details)

Save the Children, IDELA. (2023). *About IDELA*. IDELA. https://idela-network.org/about/. (details)

Seiden, J. (2021). *A Framework for Creating Short Forms of Internationally Used Direct Assessments*. GPRL/IPA Methods & Measurement meeting, Northwestern University, Evanston, IL. (details)

Shavitt, I., Ayres de Araujo Scatollin, M., Suzart Ungaretti Rossi, A., Pacífico Mercadante, M., Gamez, L., Resegue, R. M., Pisani, L., & do Rosário, M. C. (2022). Transcultural adaptation and psychometric properties of the International Development and Early Learning Assessment (IDELA) in Brazilian pre-school children. *International Journal of Educational Research Open*, *3*, 1–8. https://doi.org/10.1016/j.ijedro.2022.100138. Available from https://www.sciencedirect.com/science/article/pii/S2666374022000176. (details)

Sims, S., Anders, J., Inglis, M., Lortie-Forgues, H., Styles, B., & Weidmann, B. (2023). *Experimental education research: rethinking why, how and when to use random assignment* (CEPEO Working Paper Series 23–07). UCL Centre for Education Policy and Equalising Opportunities. https://econpapers.repec.org/paper/uclcepeow/23-07.htm. (details)

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, USA. (details)

Styles, B., & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research — methodological debates, questions, challenges. *Educational Research*, *60*(3), 255–264. https://doi.org/10.1080/00131881.2018.1500194. (details)

Sullivan, G. M. (2011). Getting off the "Gold Standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education*, *3*(3), 285–289. https://doi.org/10.4300/JGME-D-11-00147.1. Available from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3179209/. (details)

Thai, K.-P., Bang, H. J., & Li, L. (2022). Accelerating early math learning with research-based personalized learning games: A cluster randomized controlled trial. *Journal of Research on Educational Effectiveness*, *15*(1), 28–51. https://doi.org/10.1080/19345747.2021.1969710. Available from https://www.tandfonline.com/doi/full/10.1080/19345747.2021.1969710. (details)

Torgerson, C. J., & Torgerson, D. J. (2012). The Need for Randomised Controlled Trials in Educational Research. In *The Need for Randomised Controlled Trials in Educational Research* (pp. 203–214). Routledge. https://doi.org/10.4324/9780203123829-29. Available from https://www.taylorfrancis.com/chapters/edit/10.4324/9780203123829-29/need-randomised-controlled-trials-educational-research-carole-torgerson-david-torgerson. (details)

Wolf, S., Halpin, P., Yoshikawa, H., Pisani, L., Dowd, A. J., & Borisova, I. (2016). *Assessing the Construct Validity of Save the Children's International Development and Early Learning Assessment (IDELA)*. Global TIES for Children, New York University, Save the Children. (details) https://resourcecentre.savethechildren.net/document/assessing-construct-validity-save-childrens-international-development-and-early-learning/

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, *41*, 21–36. https://doi.org/10.1016/j.ecresq.2017.05.001. Available from https://www.sciencedirect.com/science/article/pii/S0885200617301357. (details)

Yousafzai, A. K., Rasheed, M. A., Rizvi, A., Shaheen, F., Ponguta, L. A., & Reyes, C. R. (2018). Effectiveness of a youth-led early childhood care and education programme in rural Pakistan: A cluster-randomised controlled trial. *PLOS ONE*, *13*(12), 1–14. https://doi.org/10.1371/journal.pone.0208335. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208335. (details)

Zhang, Z. (2022). *WebPower package - RDocumentation*. https://www.rdocumentation.org/packages/WebPower/versions/0.6. (details)

Annex A: full research timeline

The following table provides an overview of the operational and research timeline for the rollout of EIDU in the four selected sub-counties and the implementation of the RCT.

Table 11. *Full research timeline*

Date	Kenyan academic calendar	Action
Early August 2022	Half-term break / Term 2	Randomisation finalised and communicated to schools
Early September 2022	Term 2	Recruitment of enumerators
w/c 26 September 2022	Term 3	Roll out of treatment arm begins (DPL tool only)
27–30 September 2022	Term 3	Enumerator training
3–18 October 2022	Term 3	Baseline assessment
w/c 9 January 2023	Holiday	Tayari training (ECDOs and schools in treatment group receive separate training)
w/c 23 January 2023	Term 1	Roll out of the full DPL-SP model in treatment schools
3–5 & 11–12 May 2023	Holiday / Term 2	Enumerator training
15–26 May 2023	Term 2	Midline assessment
25–28 September 2023	Term 3	Enumerator training
2–17 October 2023	Term 3	Endline assessment
January 2024 onwards	Term 1 onwards	Phased rollout of the EIDU DPL-SP model to control schools begins (phased)

Annex B: full range of a priori power calculations

The full range of a priori power calculations are presented in this annex. Other resources, including the full, adapted IDELA assessment tool, consent forms and randomisation spreadsheets, are available on request from the research team.

A range of a priori power calculations were made using the WebPower package in R (wp.crt2arm command) in order to estimate the sample size required for the RCT. The full range of calculations are outlined below. Firstly, all calculations for two arms (one treatment, one control) are outlined; then, all calculations for three arms (two treatment, one control, which was the original proposal for the RCT), are outlined as a comparison point. In both tables, the calculations for an intraclass correlation (ICC) of 0.3 are presented first, followed by the comparative calculations for an ICC of 0.2.

Table 12. *Full set of power calculations*

Key			
Minimum cluster size	**0.1 effect size**	**Middle ground (smaller effect size than 0.2, without increasing sample size too much).**	
2 arms (1 treatment, 1 control)			
School cluster sample size (incl. 1 treatment arm & control)	**Learner assessment sample size per school**	**Minimum detectable effect size**	**Power**
0.3 ICC			
258	25	0.2	0.8
262	20	0.2	0.8
274	15	0.2	0.8
294	10	0.2	0.8
300	10	0.19	0.8

360	10	0.18	0.81
1170	10	0.1	0.8
1090	15	0.1	0.8
1050	20	0.1	0.8
330	20	0.18	0.8
1030	25	0.1	0.8
320	25	0.18	0.8
290	25	0.19	0.8
0.2 ICC			
184	25	0.2	0.8
190	20	0.2	0.8
200	15	0.2	0.8
220	10	0.2	0.8
3 arms (2 treatment, 1 control)			
School cluster sample size (incl. 2 treatment arms & control)	**Learner assessment sample size per school**	**Minimum detectable effect size**	**Power**
0.3 ICC			
292	25	0.2	0.8
298	20	0.2	0.8
309	15	0.2	0.8
329	10	0.2	0.8
360	10	0.19	0.8
420	10	0.18	0.81
1320	10	0.1	0.8

1230	15	0.1	0.8
1200	20	0.1	0.8
360	20	0.18	0.8
1170	25	0.1	0.8
360	25	0.18	0.8
330	25	0.19	0.81
1140	30	0.1	0.8
1110	60	0.1	0.8

0.2 ICC

207	25	0.2	0.8
214	20	0.2	0.8
226	15	0.2	0.8
249	10	0.2	0.8
991	10	0.1	0.8
897	15	0.1	0.8
850	20	0.1	0.8
821	25	0.1	0.8