Received: 27 October 2023 | Accepted: 11 October 2024

DOI: 10.1111/bjet.13533

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Supporting equitable access to learning via SMS in Kenya: Impact on engagement and learning outcomes

Katy Jordan[1] | Christina Myers[2] | Kalifa Damani[3] | Phoebe Khagame[4] | Albina Mumbi[4] | Lydia Njuguna[4]

[1]Faculty of Education, University of Cambridge, Cambridge, UK

[2]Overseas Development Institute, London, UK

[3]Jigsaw Consult, London, UK

[4]M-Shule, Nairobi, Kenya

Correspondence
Katy Jordan, Department of Educational Research, Lancaster University, Lancaster, UK.
Email: k.jordan@lancaster.ac.uk

Present address
Katy Jordan, Department of Educational Research, Lancaster University, Lancaster, UK

Funding information
EdTech Hub; United States Agency for International Development; Foreign, Commonwealth and Development Office

Abstract: The use of SMS messaging for education has grown in recent years, with particular attention recently during the Covid-19 pandemic. Mobile phones often have high levels of ownership in low-income contexts compared to computers, and lower connectivity requirements, which arguably make this a more equitable medium than data-heavy online instruction, for example. However, given that gender can be a factor to influence mobile device access and use, it is also important to consider educational applications through a gender lens, to avoid further exacerbating digital divides. In this paper, we present an analysis of server log and evaluation data in relation to a literacy-focused initiative for primary-aged learners carried out in Kenya as part of the Tusome programme and through the SMS-based M-Shule education platform, which does not require an Internet connection or smartphone to run. The extent of engagement with the platform varies according to gender and location within the country. The data also demonstrate a positive impact on learning outcomes regardless of learners' gender and location. Furthermore, the learning gains are shown to be relatively cost-effective in comparison with educational technology interventions in similar contexts. The findings show that this low-connectivity adaptive model has a positive impact on learning outcomes. It is a scalable approach to support a range of learners in Kenya, providing more support to learners who need it, and leading to increased foundational learning outcomes overall. As such, the findings will also be of highly relevant to other low-connectivity contexts.

KEYWORDS
educational technology, Kenya, literacy, mobile learning, SMS

Practitioner notes

What is already known about this topic

- Mobile phones can be used as a means to support learning, through mobile learning and SMS, particularly in low-connectivity contexts, although there is a lack of rigorous evidence of impact upon learning outcomes.
- Mobile phone device ownership tends to be higher than computer or wired Internet connections in many low-income contexts.
- Software applications which adapt to the learners' level have shown good potential for gender-equitable learning outcomes in low-income contexts; however, these often require an Internet connection in addition to computers or tablets to be run on.

What this paper adds

- There is a lack of contextually relevant evidence of the impact of SMS-based mobile learning applications in low-resource and low-connectivity contexts upon learning outcomes.
- Through analysis of data generated via an experimental design, this study provides evidence that literacy materials delivered through an SMS-based educational platform—M-Shule—can have a positive impact upon learning outcomes.
- Furthermore, gains are equitable in terms of learners' gender, and location, within Kenya.

Implications for practice and/or policy

- Mobile phones can be an effective way of reaching learners to provide additional educational support as part of existing education programmes in low-connectivity environments.
- Learning gains using M-Shule are evidenced as significant and relatively cost-effective.
- Existing high-quality learning materials developed in other media can be effectively adapted to SMS to reach learners particularly who are out-of-school or during periods of educational disruption.

INTRODUCTION

The greater levels of ownership of mobile phones, compared to wired Internet connections and computers, has long been regarded as a particular advantage of the potential for 'mobile learning' to be used in Low- and Middle-Income Countries (LMICs) (Sharples & Pea, 2014; Wagner, 2014). In Kenya, households are much more likely to have mobile phones—often 'feature phones' rather than smartphones—than computer hardware and wired Internet connections. For example, mobile phone subscription rates far outweigh Internet use (117% of the population—suggesting that ovit is common to have more than one mobile phone per person—compared to 33%; DataReportal, 2023).

Sending messages easily and cheaply using mobile phones—either as SMS or using apps such as WhatsApp—is a particular advantage of mobile learning in LMICs. Examples can be found at multiple levels of education systems. For example, mobile learning may allow students to directly engage with educational materials and activities. Mobile messaging has also been used to support parents and caregivers in helping their children with education at home, and networks to support teachers and their professional development (Jordan, 2023). However, while there is good potential for the use of m-learning in low-resource contexts, there is a lack of contextually appropriate research and evidence for the impact upon learning outcomes (Jordan, 2023).

The use of mobile messaging is a topic which has received intense focus as a result of the Covid-19 pandemic, and there is some evidence to suggest that applications have been effective in supporting education at a distance. As the disruption caused by Covid-19 subsides, it is now timely to review experiences and evidence in relation to the use of messaging-based educational initiatives, in order to look ahead to future preparedness and how any benefits can be incorporated into education in the longer term.

SMS as a platform for education

Mobile phones have been used to support continued educational provision during periods of school closure in two main ways; to 'nudge' caregivers in helping with educational activities at home, or as a medium for delivering educational materials and assessments directly to learners (Jordan, 2023). During periods of school closure due to Covid-19, the use of mobile phone-based messaging to support caregivers to study with their children at home has received a greater focus (eg, Roberts et al., 2021; Wolf et al., 2021), and results have been mixed (Jordan et al., 2023).

While fewer studies have focused on the latter use—to directly support students—this mode may be more useful to focus upon as education returns to in-person schooling. A key example is the recent study undertaken in Botswana, which evaluated interventions using SMS messaging to improve children's numeracy learning outcomes through the implementation of a randomised control trial (RCT) involving 4500 households (Angrist et al., 2020a, 2020b). The first intervention included SMS messaging with numeracy exercises, while the second intervention also provided support via phone calls. Both interventions led to significant learning gains relative to a control group. Of the two treatment arms, greater learning gains were reported for the intervention that included SMS messaging and phone-calls compared to the arm which relied exclusively on SMS messaging. However, the additional phone calls made the gains less cost-effective. It is also notable that the learning gains with SMS messaging were not sustained over time, and the pedagogic design also made a significant difference, only being effective when using targeted instruction (Angrist et al., 2020b).

Although the Covid-19 pandemic has fostered interest in SMS messaging for educational purposes in LMICs, there are already a number of SMS-based education platforms in operation, although little empirical research has been published from current initiatives and companies. A notable exception, however, is the 'Shupavu291' product from Eneza Education. Operating in Kenya, Ghana and Cote D'Ivoire, Shupavu291 is a mobile phone-based educational platform, which provides learners with curriculum-linked educational materials, quizzes, and allows users to submit questions to teachers, via SMS (Kizilcec & Goldfarb, 2019).

Data mining techniques have been applied through a group of studies seeking to identify trends in a large-scale dataset of Kenyan Shupavu291 users (n = up to ~93,000). Kizilcec and Goldfarb (2019) examined which factors are associated with higher quiz scores, and found that a growth mindset, gender (higher quiz scores associated with female students), higher school grades, and greater satisfaction with the learning environment play a role. The study also found that receiving study assistance may help, but is dependent on who provides it; students who were given study help from their parents or a tuition teacher had lower quiz scores than those who received help from friends or classmates, which may suggest that student–student communication through messaging could be beneficial (Kizilcec & Goldfarb, 2019, p. 5). However, considering the likelihood of learners correctly answering quiz questions over time, the authors did not find evidence to suggest that use of the platform is associated with enhanced learning gains (Kizilcec & Chen, 2020).

Another significant finding from the Shupavu291 data analysis relates to patterns of learner persistence, including disengagement and re-engagement. A large proportion of learners did not use features of the platform after initial enthusiasm (Kizilcec & Chen, 2020). To an extent, this shows similarity to patterns of persistence in other forms of digital education where engagement has been shown to follow a heavy-tailed distribution over time, such as MOOCs (Jordan, 2015; Kizilcec et al., 2013). However, the wider literature on persistence is focused towards higher education and high-connectivity online learning. There is a lack of other studies of persistence in the context of SMS-based applications with school-aged learners, and the Shupavu291 data suggest that there are also key differences compared to other forms. As Shupavu291 content is aligned to the curriculum, use varies according to the school year; higher levels of activity are associated with self-directed study in school holidays and in preparation for examinations. Only minor differences in engagement were observed according to grade level (Chen & Kizilcec, 2020). Levels of use were also analysed from (pre-Covid) periods of disruption to in-school education, with the analysis in Kenya focused upon data during periods of school closure in 2017 due to political unrest. Platform use increased during disruption, and preferences for types of content accessed changed (eg, greater demand for course content, rather than quizzes) (Kizilcec et al., 2021). While there is potential for this form of EdTech to support education in a flexible way to suit learners' demand, there is also an open question about how engagement can be sustained in the longer term.

Gender, equity and EdTech

In this project, we also place particular focus upon considering whether the impacts of using EdTech are equitable in terms of gender; that is, whether any benefits are experienced to an equal extent for girls and boys. This is for various reasons, depending on the context. Cultural barriers and other forms of structural inequalities may mean that access to mobile phones for learning may not be equitable. Girls are often less likely to be allowed access than boys (Crompton et al., 2021; Watson et al., 2023; Zubairi et al., 2021), but this is not always the case (Gupta et al., 2023). There can also be stark differences in terms of which caregiver allows their phone to be used by children, which may tend to be female caregivers (Gupta et al., 2023). However, girls tend to outperform boys in terms of early literacy, while boys on average have a slight advantage in terms of maths (Fonseca et al., 2023). The nature of gender gaps is complex and it is therefore important to consider the relationship between EdTech and gender carefully and in context. Yet, relatively few EdTech studies disaggregate findings by gender (Jordan & Myers, 2022). Access to education may also vary according to location, with lower school attendance in counties in Arid and Semi Arid Lands (ASAL) (UNICEF Kenya, 2018).

Software which adapts to the level of the learner (facilitating 'personalised learning') has demonstrable positive impacts for learners in LMICs (Major et al., 2021). While 'personalised learning' is a wide-ranging term and encompasses a variety of different technical approaches (Van Schoors et al., 2021), simple adaptation to the learners' level can be particularly useful in LMIC contexts, reflecting the pedagogical approach of 'teaching at the right level' (Angrist et al., 2023). Jordan and Myers (2022) compared to the relative impact on learning outcomes for girls from a range of EdTech interventions undertaken in LMIC contexts, which suggested that software which adapts to the learners' level may be particularly beneficial. A key recent example is a study in Malawi, which showed that a digital personalised learning application delivered using tablet computers could mitigate gender learning gaps in contexts where girls are hindered from acquiring numeracy skills at the same rate as boys (Pitchford et al., 2019). However, personalised learning software frequently relies upon computers, tablets or smartphone devices and may not be accessible in contexts where levels of computer and tablet ownership may be much lower than mobile devices (which are often 'feature phones' rather than higher connectivity smartphones), as is the case in the context of Kenya (Cotter Otieno & Taddese, 2020; DataReportal, 2023). As a platform which uses learners' quiz scores to adapt their progress, the present study also makes a contribution to this research gap.

Study context and research questions

This study focuses upon a platform called M-Shule, which uses SMS to deliver educational content without requiring an Internet connection. Created in 2017 and meaning 'mobile school' in Kiswahili, M-Shule is an EdTech platform that learners in Kenya engage with through mobile phones. To date, the platform has reached approximately 20,000 households from 30 Kenyan counties, mainly to improve primary school children's foundational learning (UNESCO Institute for Lifelong Learning, 2022). M-Shule uses SMS in two main ways: to deliver educational content, and quizzes. Learners' performance in the quizzes determines their progress to the next unit of content, so to an extent the content of the SMS is matched with the children's needs and learning level. Through this mechanism, children receive harder questions and lessons as they gain mastery and easier content if they need to further build foundational or remedial skills.

In this study, we draw upon data about the use of the platform in the wake of school closures in Kenya prompted by the Covid-19 pandemic (Ngware & Ochieng, 2020; Tembey et al., 2021), focusing on an intervention based on adapting materials from the Tusome programme into a format suitable for remote delivery by SMS. The Tusome programme began in 2015 and includes instructional materials and the use of tablets to enhance teacher coaching and oversight (Piper et al., 2018). The programme focuses on enhancing the quality of Kenyan early grade literacy by improving teacher quality and has been leveraging technology to support this goal. The programme targeted Grades 1 to 3 in all 23,000 Kenyan public primary schools and implemented multiple approaches to improve children's literacy outcomes (Piper et al., 2018). While the Tusome programme was a broader educational intervention, implemented at several levels of the educational system, the particular element which was adapted for delivery through SMS was the content aimed at learners via books.

Following the disruption caused by Covid-19 school closures, the M-Shule platform worked with the Tusome programme, to digitise and adapt the learners' content of the Tusome programme to delivery to be delivered to learners through the M-Shule platform via SMS as a booster programme. Sixty interactive literacy lessons were developed, including 30 each in English and Kiswahili, comprising 10 per grade for children in Grade 1 to 3 (see Appendix). Screenshots showing examples of the delivery of content and responses via

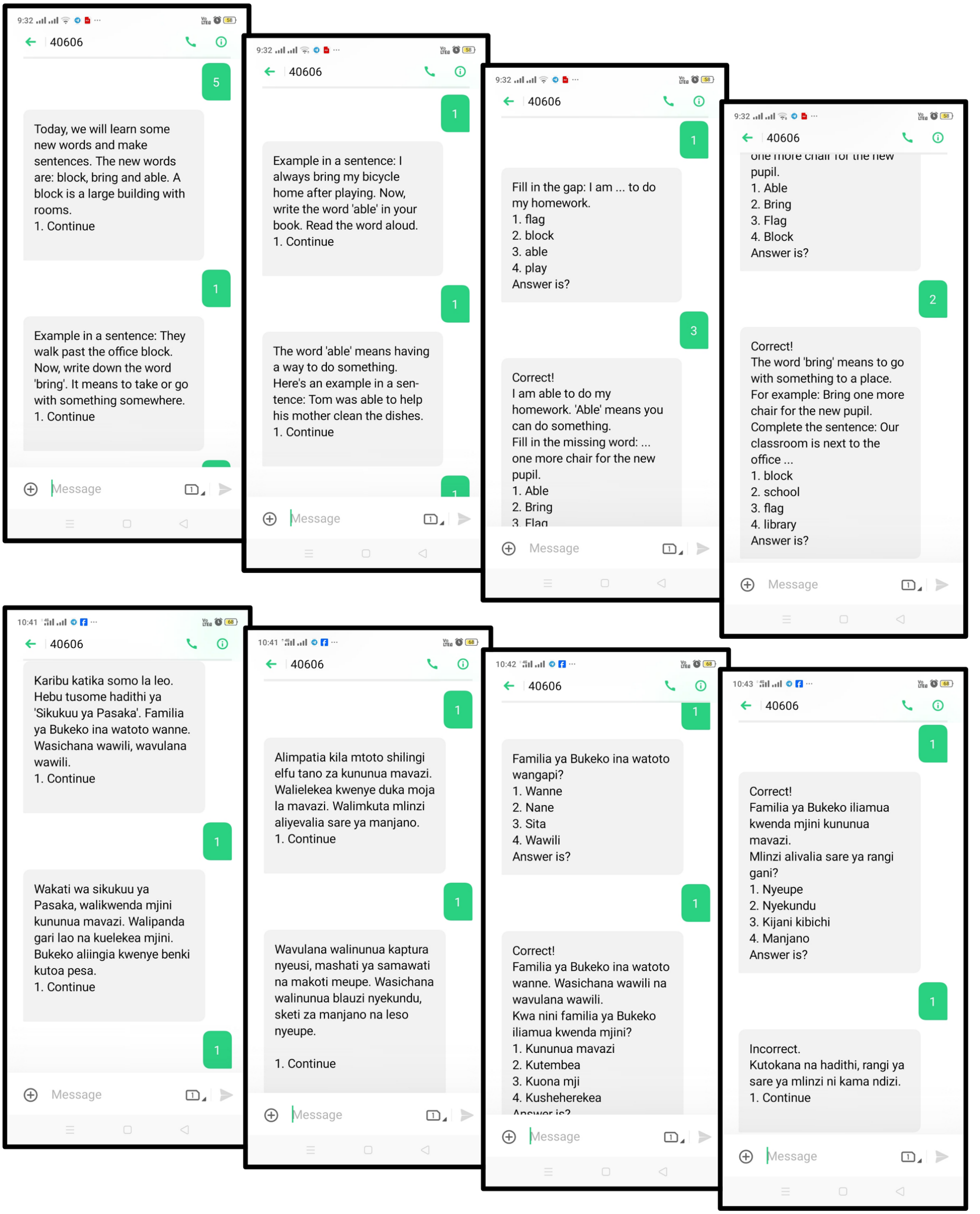

FIGURE 1 Examples of screenshots to show delivery of the Tusome content via SMS, in English (top row) and Kiswahili (bottom row).

SMS are shown in Figure 1. Children from marginalised and rural areas were prioritised as participants. Parents and caregivers were also sent supportive messages to engage with the literacy lessons and to use the M-Shule platform.

To evaluate the effectiveness of adapting Tusome literacy lesson content for learners for delivery by SMS through the M-Shule platform, an experimental study was undertaken with a group of learners. This was carried out as schools reopened in Kenya. Learners from the existing user base were assigned to treatment or control groups. During the experimental period, learners in the treatment group had access to the new Tusome literacy lessons via SMS, while those in the control group did not (see Section 'Assessment data' for more detail).

The overall goal of this study was to explore the use of the Tusome lessons delivered through the M-Shule platform in terms of learning gains and addressing inequalities. The study was guided by the following research questions:

1. Are gender, location and treatment group associated with any learning gains from using the literacy lessons via the M-Shule platform?
2. Are gender and location associated with different levels of use of the literacy lessons via the M-Shule platform?

To address the research questions, we draw upon two forms of quantitative data related to the use of the educational materials through the M-Shule platform. This includes server log data about the extent of engagement for all learners who used the content ($n=829$), and a sub-group of learners ($n=94$) who were part of an experimental trial to assess the learning gains associated with use of the platform relative to a control group who did not access the content ($n=85$). The data were analysed using parametric and nonparametric statistical tests, as appropriate for the distribution and characteristics of the different forms of data collected.

METHODS

The two research questions were addressed through two different ways of sampling and analysing learners' use of the M-Shule materials. Research question 2 utilised server log data—that is, automatically generated observational digital data about the level of use of the platform by all users. This dataset comprised all the learners who had accessed any of the materials ($n=829$).

Examining research question one relied on comparing two different groups of users of the M-Shule, one of which received access to the new Tusome literacy lessons intervention through the platform (treatment, $n=94$) or not (control, $n=85$), to explore whether any learning gains occurred. It therefore required an experimental approach to secondary data analysis of assessment data. The 94 learners within the treatment group are a sub-sample of all the learners who accessed the materials, so are also present in the server log dataset used to address research question 2. The 85 learners in the control group were also drawn from the user base of M-Shule, but were not able to access the new materials prior to the endline assessment.

As answering the research questions involved utilising different approaches and types of data, a mixed methods approach was used as the grounding methodology for this study (Creswell & Plano Clark, 2017). The characteristics of both sets of data used (assessment data and server log data), as well as how they were analysed to answer each research question, are discussed in the following sections.

Assessment data

Secondary data analysis was undertaken to answer research question one: Are gender, location and treatment group associated with any learning gains from using the literacy lessons via the M-Shule platform? A dataset of learning outcome data recorded from a sample of users of the platform was utilised. Learning assessments were administered by telephone (Angrist et al., 2020c) to a sample of participants at two points during the intervention.

The sample included a treatment ($n=94$) and control ($n=85$) group. Potential participants were drawn from a pool of users who had previously registered with the system and may have used other applications in the past. As the intervention materials were targeted at learners in Grades 1 to 3, learners' ages ranged from 5 to 11 years. Initially, 500 students were sampled for this evaluation; 250 students were randomly assigned to the control group and 250 to the treatment group. However, the sample included telephone numbers which were no longer active; contacts which did not respond after three attempts (198—37%—from the original sample of 500) were considered inactive and a further 1000 contacts (500 randomly assigned to the treatment group, 500 to the control group) were sampled. Learners in the treatment group received the new Tusome literacy lessons SMS intervention through M-Shule, while those in the control group did not receive access to the new content (or encouragement to make use of any previous content they may have accessed previously through M-Shule). Note that at the time of the intervention, schools in Kenya had fully reopened (Mukoya, 2021). It is assumed therefore that learners in both the treatment and control groups were attending school, but those in the treatment group were also receiving the Tusome programme through M-Shule as an additional educational support.

Not all of the students invited to take part participated in the assessments. It is not clear exactly why students did not take part in the assessment; to an extent, being unavailable when the telephone calls for the assessment were made would have been a likely factor to an extent. The study was also undertaken during the pandemic, which would have placed further pressures on potential participants' time. A total of 244 students were assessed at baseline in September 2021, and 179 at endline in November/December 2021 (26% attrition rate). While the attrition rates are not ideal and reduce external validity (Murnane & Willett, 2011), they are not unusually high given the context of telephone assessment during the pandemic (Rodriguez-Segura & Schueler, 2022). In terms of internal validity, attrition rates were slightly higher in the treatment group (24% for control and 29% for treatment groups). Balance checks were undertaken during the primary data collection at baseline and endline in relation to several factors (including school grade, caregiver education level, household size and income) and were satisfied.

The ASER reading scale was used for the assessments (Vagh, 2009), and the assessments were carried out in English. The ASER tool is an ordinal scale of increasingly challenging tasks; the enumerator records the highest level achieved by the student. Note that when the data were collected, two different versions of the scale were used—a five point scale at baseline, and a six point version at endline. The six point scale was used at endline as it was observed during initial data collection that it would be useful to have greater variation within the dataset, and the sixth point was used at endline to be able to distinguish further. Consequently, comparisons cannot be conclusively drawn between an individuals' performance at baseline and endline which reduces the validity of the study to an extent. Instead, we focus here upon the endline assessment alone, drawing comparisons between the treatment and control groups at this point. The analysis therefore follows a single difference impact estimate causal inference research design (White & Sabarwal, 2014), with the additional benefit of having comparability across treatment and control groups at baseline, but not being in a position to calculate exact rates of change for individual learners across the course of the intervention.

Examples of the questions asked to assess which of the scale points corresponded to the learners' level are shown in Table 1. The assessments were carried out in English. The scale points were recorded as follows: 1, beginner; 2, letter; 3, word; 4, paragraph; 5, story; and 6, comprehension. As the scale generates nonparametric (ordinal) data (Field, 2009), an ordinal logistic regression was utilised, using SPSS, to answer research question one. The endline ASER test scores were assigned as the dependent variable, with binary categorical variables of treatment group (control or treatment), child's gender (male or female) and location (ASAL or non-ASAL county) as factors, and the child's school grade as a

TABLE 1 Examples of questions used when administering the ASER scale assessments by telephone.

Scale point	Category	Example
1	Beginner	[learner does not meet criteria for Letter level]
2	Letter	Next, some letters will appear on the screen. Ask the child to read any 5 letters. If they do not choose, point out 5 letters to them. e d w s c g h z i q … Did the child identify AT LEAST 4 letters correctly?
3	Word	Next, some words will appear on the screen. Ask the child to read any 5 words. If they do not choose, point out 5 words to them. hand star bus cat book day few old sing bold. … Did the child identify AT LEAST 4 words correctly?
4	Paragraph	Next, some sentences will appear on the screen. Ask the child to read them aloud. Juma likes his school. His class is in a big room. Juma has a bag and a book. He also has a pen. … Did the child read with fluency and/or without many mistakes?
5	Story	In the next 2 messages, a story will appear on your screen. Ask your child to read it aloud. A big tree stood in a garden. It was alone and lonely. One day a bird came and sat on it. The bird held a seed in its beak. … It dropped the seed near the tree. A small plant grew there. Soon there was another tree. The big tree was happy. … Did the child read with fluency and/or without many mistakes?

covariate. Given the focus of the research questions, gender and location were included as key variables of interest in the analysis. Gender and location were self-reported by participants during the telephone interviews. Counties were categorised according to whether they are predominantly ASAL or non-ASAL (Birch, 2018) as although primary school enrolments are now high for Kenya overall, there is considerable variation and substantially lower school attendance in ASAL areas (UNICEF Kenya, 2018). The results of the ordinal logistic regression are presented in Section 'Learning gains', in accordance with APA guidelines.

Although it was not possible to measure the gains for individual students, the baseline data was useful to confirm that there were no significant differences between the groups initially. A similar ordinal regression, with treatment, gender and location as factors and grade as a covariate, showed no significant differences according to treatment group at the baseline stage (full results for this analysis are shown in Table A4 in the Appendix).

In addition to using the data on learning outcomes to examine whether use of the system leads to a demonstrable gain relative to a control group, we also consider to what extent these gains are cost-effective. To do so, we follow the approach used by Angrist et al. (2020a), which also used a version of the ASER test. This approach is also useful as comparisons are made here between control and treatment at endline; the inconsistency of the scales used at baseline and endline prevent the use of a cost-effectiveness metric which relies on the rate of gain (such as Learning Adjusted Years of Schooling, LAYs; Evans & Yuan, 2019). This involves expressing the difference at endline between the treatment and control group in terms of standard deviations within the control data. By combining this with information about (i) the total cost of implementation and (ii) number of learners, an estimate of cost per standard deviation gain can be calculated.

Server log data

Interactions between learners and the M-Shule platform are captured in server log data. The log was exported for analysis as a CSV file. Note that no names were included when exported, and the data were held in secure password-protected digital storage for analysis.

The raw dataset comprised a total of 64,721 interactions with the system (one per row), from 829 learners. Interactions are any SMS responses and included any individual content items which a learner accessed within a subtopic (if the same item was accessed multiple times, each would be logged separately) and attempts at quiz questions (similarly, each attempt would be a separate interaction). Interaction figures are a combination of different actions but overall, this figure provides an approximation of the extent to which learners used the platform overall. Each interaction had some information about the learner associated with it (such as a unique identifier, gender, class level and county) in addition to the course and topic of that activity, date and a proficiency score (for quizzes, typically at the end of a sub-topic).

The content is divided into three class levels, each comprising an English booster and a Kiswahili booster course (six courses in total). Content within each course was divided into topics and sub-topics. Learners could choose to study as many or few sub-topics as they would like. A full list of the topics, subtopics and number of unique learners for each is shown in the Appendix (Tables A1–A3).

Similar to the learning outcome data, gender and location (ASAL or non-ASAL) were the explanatory variables included in the analysis. Measures of engagement with the system included (i) the number of sub-topics learners interacted with, (ii) the number of times they interacted with the system (this included every content unit or quiz attempt), and (iii) quiz scores logged. Similar to the learning gains data, the analytical approach used was a regression analysis. The measures here are scale variables, but differ in their distribution of values. Final quiz scores followed a normal distribution, however, the number of sub-topics and interactions did not. Log transformation was applied to the number of interactions, which then allowed analysis through a univariate ANOVA test. However, transformation of the distribution of number of sub-topics was not successful in achieving a normal distribution, so as a result, nonparametric tests were applied to this measure instead (Kruskal–Wallis tests).

Ethics

In the design and undertaking of this research project, we followed and applied the British Educational Research Association (BERA) ethical guidelines (BERA, 2018). For the secondary analysis of preexisting data (learning evaluation and server logs), the collection of data for analysis had been part of the terms and conditions of platform use. Names and personal identifiers were removed from the datasets before being shared with the team for analysis. An application was made to the EdTech Hub ethics panel, and approval was granted before commencing work with the datasets. Furthermore, an application was made to the ethical review board of Maseno University, Kenya, and to NACOSTI for a research permit (Laterite, 2021).

RESULTS AND DISCUSSION

Learning gains

An ordinal regression was carried out to investigate whether gender and location were associated with differences in endline scores. Learners' school-grade level was also included in the model as a covariate. While there was no significant relationship between endline scores and gender and location, the model confirmed a significant relationship between treatment or control group and endline scores.

The deviance goodness-of-fit test indicated that the model was a good fit to the observed data, $\chi^2(131) = 100.675$, $p = 0.977$, but approximately half of the cells were sparse with zero frequencies in 51.8% of cells. However, the model statistically significantly predicted the dependent variable over and above the intercept-only model, $\chi^2(4) = 16.301$, $p < 0.001$.

The odds of learners in the treatment group outperforming their counterparts in the control group was 2.677, 95% CI [1.521–4.712], $\chi^2(1) = 11.658$, $p = 0.001$.

Table 2 shows the results of the regression analysis in full. The differences between the treatment and control group are also clear when shown visually, in Figure 2. While both groups show a modal category of responses at the level of 'comprehension', there is notably lower variation in the treatment group (grey bars) compared to the control (white bars), which suggests that the intervention was effective.

In order to consider the relative cost-effectiveness of the learning gains, we follow the approach used by Angrist et al. (2020a), which was also applied to ASER scale data. The mean average endline score for the control group is 4.56, with a standard deviation of 1.38. The mean average endline score for the treatment group is 5.20; therefore, the treatment effect is 0.64, which translates into a 0.46 standard deviation gain. The fixed costs included the cost of staff time for technical development, the adaptation of Tusome content to the platform, quality assurance and evaluation. Given the costs, average learning gain and the total number of learners reached (the total of 829 learners who interacted with the platform), this translates into an average $22.70 USD cost per standard deviation gain, which is relatively cost-effective in comparison with benchmarks (Angrist et al., 2020a). However, it is also important to note that the initiative being implemented here is relatively small-scale, and that following the initial costs associated with adaptation of the content, cost-effectiveness would be enhanced if implemented at a larger scale. Despite this, it is useful to consider cost-effectiveness in relation to the cost of the implementation, as this is an aspect of educational technology evaluations which is rarely considered but may be critical for policymakers facing budgetary decisions (Nicolai et al., 2023).

Engagement with the platform

The time period covered by the server log data runs from 15 December 2020 to 27 February 2022. The number of interactions per day during this period is shown in Figure 3, which shows that there were two periods of more intensive activity during this time: January to April

TABLE 2 Results of the regression analysis.

Predictor	Estimate	*SE*	Wald	*p*	Exp(*b*)	95% CI
Coefficients						
Grade	0.370	0.157	5.529	0.019	1.447	1.063–1.969
Gender	−0.311	0.283	1.202	0.273	0.733	0.421–1.277
Location	−0.199	0.285	0.486	0.486	0.820	0.469–1.433
Treatment	0.985	0.288	11.658	0.001	2.677	1.521–4.712
Intercepts						
1. Letter	−4.006	1.121	12.766	0.000	0.018	
2. Word	−2.160	0.650	11.043	0.001	0.115	
3. Paragraph	−0.370	0.549	0.454	0.500	0.691	
4. Story	0.935	0.549	2.902	0.088	2.547	
5. Comprehension	1.534	0.556	7.604	0.006	4.638	

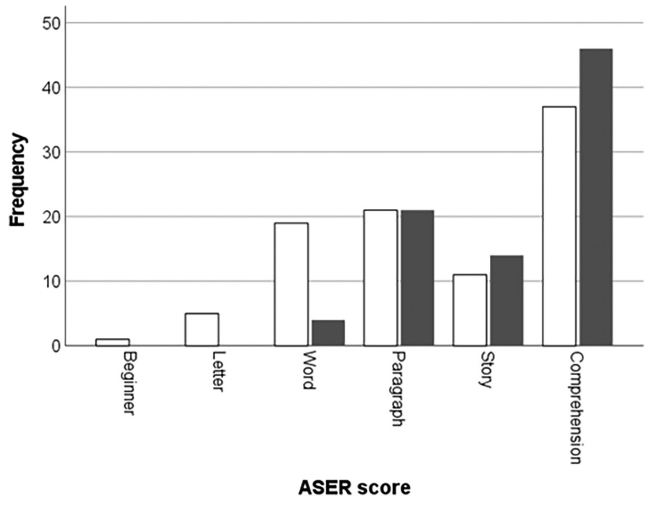

FIGURE 2 Frequency of different ASER scores in the endline data collection. White bars represent the control group, and grey bars the intervention group.

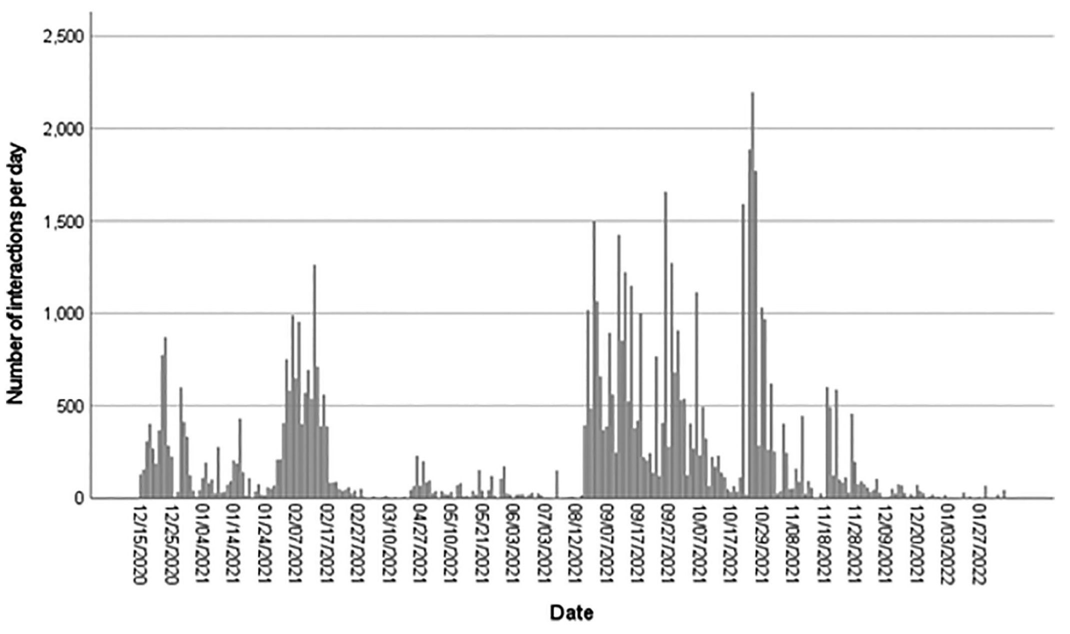

FIGURE 3 Bar chart showing frequency of interactions with the platform over time.

2021, and September to December 2021 (the latter reflecting the time period of the learning assessments). Note that school closures were initiated in March 2020 (Cotter Otieno & Taddese, 2020) and fully reopened in January 2021 following a staggered reopening (Mukoya, 2021). The time period shown in Figure 3 therefore starts with the reopening of schools, and the initiative serves to give learners a boost following a period of learning loss rather than substitute for in-person schooling.

Where gender data were available, the server log dataset was approximately evenly distributed. However, for 154 learners—18.6% of the dataset—gender data were not included. Of the remaining 675 learners, 50.7% were recorded as female and 49.3% as male.

Consistent with the approach used in relation to analysis of learning outcome, learners' locations were re-coded into a binary categorisation of ASAL or non-ASAL counties. In terms of location, the dataset is skewed slightly towards learners in ASAL counties, with 39.0% of learners based in an ASAL county, while 27.5% gave their location as a non-ASAL county. However, a substantial proportion—278 learners—were missing data about their location. Considering only the learners who logged a final score, 39.7% were located in an ASAL county compared to 26.1% in a non-ASAL county (with 161 missing data). Like gender, this does not suggest a difference in likelihood of completing topics according to learners' location.

In order to examine whether there were overall differences in level of use of the platform, we considered three different metrics: the number of sub-topics which learners had interacted with (logging at least one interaction); the total number of interactions logged per learner; and the final scores logged by learners who completed sub-topics.

Number of sub-topics studied

The first two analyses drew upon the full dataset of all learners who had interacted with the system (829 learners). A full breakdown of the number of unique learners who engaged with each sub-topic, according to course, is shown in the Appendix and illustrated graphically in Figure 4 (English courses) and Figure 5 (Kiswahili courses). While demand for English courses appears to be higher than for Kiswahili, both sets of curves show that a relatively small proportion of learners persisted beyond short-term engagement, which reflects trends in other studies (eg, Kizilcec & Chen, 2020).

Given the focus of the study, we considered whether gender and location are factors associated with the number of sub-topics a learner engages with. The data did not meet the normality assumption for conducting an ANOVA, even after data transformation manipulation. Kruskal–Wallis tests were therefore performed. The mean number of sub-topics completed by male (3.78) and female (3.84) students did not differ significantly (H (1, $n=545$) $=0.0929$, $p=0.761$). There was also no significant difference (H (1, $n=545$) $=1.117$, $p=0.291$) in the number of sub-topics completed by students in ASAL (3.95) and non-ASAL (3.60) counties. While girls in ASAL locations studied a greater number of sub-topics overall (4.22 compared to 3.22 for girls in non-ASAL locations, or 3.65 and 3.94 for boys respectively), no significant difference was found between the groups (H (3, $n=545$) $=1.88$, $p=0.598$).

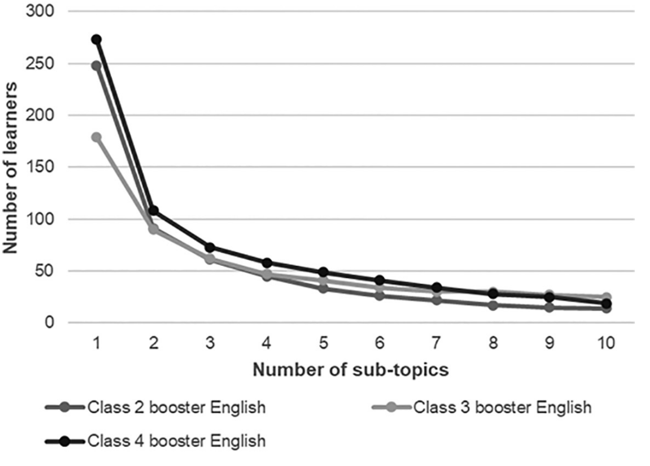

FIGURE 4 Frequency of learners by total number of sub-topics completed, in English.

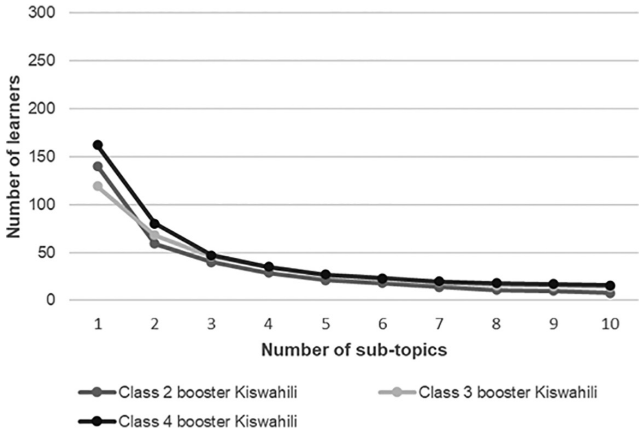

FIGURE 5 Frequency of learners by total number of sub-topics completed, in Kiswahili.

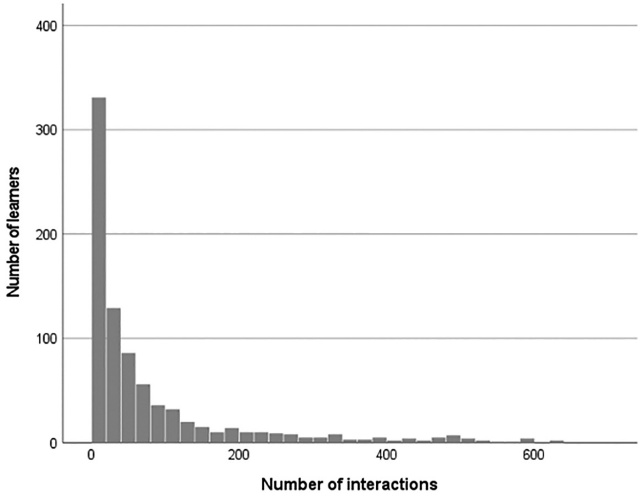

FIGURE 6 Histogram of number of interactions with the platform logged per learner.

Number of interactions with the platform

Similarly, as observed with the number of sub-topics, the number of interactions learners had with the platform follows a steeply unequal distribution (Figure 6). A log transformation was therefore conducted on the data, prior to running an ANOVA, so that the assumption of normality could be met (Figure 7).

A univariate ANOVA was carried out to examine whether gender, location or an interaction between the two are significant predictors of number of interactions with the platform. Although on average ASAL girls logged the highest mean number of interactions (94.15, compared to 75.48 for ASAL boys or 64.88 and 85.19 for non-ASAL girls and boys respectively), no significant differences in terms of the total number of interactions with the platform according to gender (F (1) = 0.000, p = 0.990), location (F(1) = 1.516, p = 0.219) or an interaction between the two (F(1, 541) = 1.876, p = 0.171) were found.

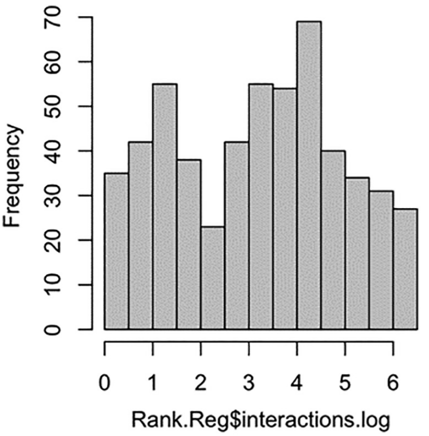

FIGURE 7 Distribution of number of interactions with the platform logged per learner following log transformation of the data.

End of sub-topic test scores

We also examined whether differences related to gender exist in terms of final test scores, for the 471 learners who logged a final test score for at least one sub-topic. While a final score is not a validated measure of learning outcomes or gain, it is a useful metric to consider for two reasons. First, logging a final score represents completion of a sub-topic. Second, scores play a technical role in progression through the system; higher scores will allow users to progress through material faster.

A final score is logged for 471 learners, which suggests that just over half of them (56.8%) studied at least one sub-topic to completion. The range of subtopics studied are shown in full in the Appendix. It is not clear from the data why the proportion of learners who engaged with the content and logged at least one final score is not higher. However, it has been noted in similar contexts that there is typically an initial fall in engagement in the use of informal mobile learning applications (a similar trend was seen with the Shupavu291 data; Kizilcec & Chen, 2020). Further qualitative research would be useful to understand the reasons behind this. In relation to the learners who logged a final score, the same proportions are approximately maintained as in the dataset overall, with 49.4% female (excluding 78 who were missing data from the total). This suggests that gender was not a factor in relation to the likelihood of learners studying topics to completion or not. The final score data recorded in the system ranges from zero to approximately 200, and is normally distributed with a mean score of 88 (Figure 8). As there could be several final scores recorded for a single module (learners could re-take questions), the highest score was used per learner, per module.

A univariate ANOVA was conducted to examine whether gender or location (ASAL or non-ASAL counties) were associated with differences in final scores logged in the platform. As learners were able to choose how many topics to engage with, language (English or Kiswahili) and course were included in the analysis as covariates. While the effects of gender and location were not significant individually, there was a statistically significant interaction between the effects of gender and location on final scores (F (1, 16,216) = 8.669, p = 0.003). While boys' scores were similar across all locations, female students from non-ASAL locations scored higher than boys and girls from ASAL locations (mean average 90.7 for girls in ASAL counties compared to 97.3 in non-ASAL counties, and 92.2 and 91.9 for boys respectively).

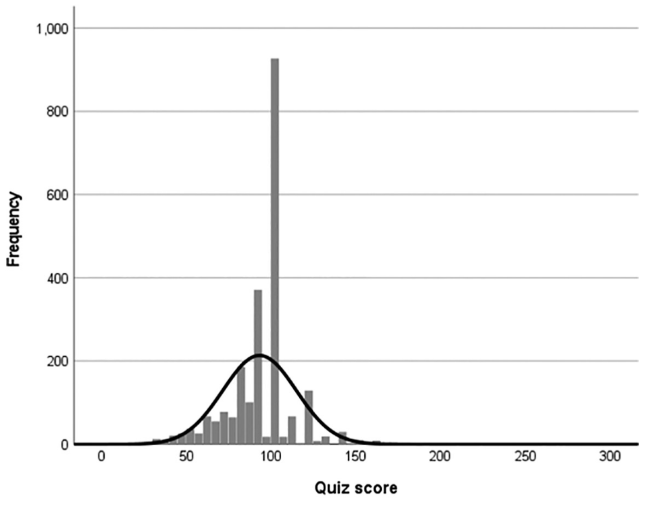

FIGURE 8 Histogram of final scores logged, with fitted normal curve.

CONCLUSIONS

In this paper, we have drawn upon data including learning outcomes and server logs to understand the use of an SMS-based education platform—M-Shule—to adaptively deliver learning content to learners in Kenya, which was integrated as part of the Tusome programme. The study makes a novel contribution in that it demonstrates the impact upon learning outcomes as a result of using an adaptive, SMS-based application, which is appropriate for a low-connectivity context. Few existing studies of mobile phone-based initiatives in low-connectivity contexts have measured learning outcomes, while adaptive software evaluations typically require additional hardware. We first sought to understand the effectiveness in terms of learning outcomes, which were shown to be significant and cost-effective in terms of standard deviation gain per dollar, relative to other educational technology initiatives in low-resource contexts (cf. Angrist, Bergman, Brewster, & Matsheng, 2020). Cost-effectiveness is a key issue for policymakers and education decision makers, presenting a key and needed contribution to advance understanding related to cost-effectiveness of personalised adaptive learning (Global Education Evidence Advisory Panel, 2023). Furthermore, we examine the learning gains and patterns in use of the content specifically through a lens of differences in use in relation to gender and socio-economic inequalities. Overall, we find that there are some differences in the ways and extent of interaction between learners and the platform in terms of these factors. The data reveal a nuanced and encouraging snapshot of the platforms' use.

The first research question was concerned with the impact of using the platform upon learning outcomes and asked '*Are gender, location and treatment group associated with any learning gains from using the literacy lessons via the M-Shule platform?*'. The endline ASER score data suggest that significant gains in literacy learning outcomes were made in the treatment group compared to the control group. As the ASER data are ordinal, it is not appropriate to use parametric approaches to calculate an effect size, but the odds ratio can be utilised as a nonparametric alternative (Leech & Onwuegbuzie, 2002) and converted to be comparable to Cohen's *d* (Sanchez-Meca et al., 2003), which yielded a value of 0.24. This suggests that the intervention performs well relative to other educational interventions in low- and middle-income country contexts (Evans & Popova, 2015). It is particularly notable compared to personalised learning applications in this context (average effect size 0.18; Major et al., 2021), with the additional advantage of not requiring a computer or laptop.

Furthermore, there were no statistically significant differences in the extent of gains according to gender, location or an interaction of the two factors. This suggests that although there are differences in terms of individual learners' engagement with the platform, the overall impact on learning outcomes is positive and equitable, when considering gender and geography. While we acknowledge that there are other forms of inequity, this is a positive sign. It would be valuable for future research to examine this further, considering socio-economic inequality in more detail, and the experiences of students with special educational needs and disability, for example.

The second research question built upon the first, to look more closely at how learners use the platform and examine '*Are gender and/or location associated with different levels of use of the literacy lessons via the M-Shule platform?*' Levels of use considered included the number of topics a learner engaged with (defined as interacting with the content of a particular topic at least once), the total number of interactions logged per learner, how likely learners were to study topics to completion, and if so the scores logged for final quizzes. In terms of both the number of topics engaged with and total number of interactions with the system, girls in ASAL counties interacted with the platform more than other groups on average. The differences may suggest that there may be a particular interest in the platform for girls in ASAL regions, but were not statistically significant. Nonetheless, it is positive to see that there were no groups which were significantly lower than the others, which suggests that this is a medium which can be adopted across all settings.

In terms of persistence and completion, gender did not appear to play a role in persistence during the courses, with the gender split in logging a final score also reflecting similar figures for those who signed up but did not complete modules. Quiz scores—logged at the end of a subtopic—was the only measure to show statistically significant differences according to gender and/or location. While boys' quiz scores are remarkably similar across both ASAL and non-ASAL counties, a sharp difference was observed between girls in different locations, with girls in ASAL areas logging the lowest scores on average, and girls in non-ASAL areas logging the highest—higher on average than boys in either group. This is interesting, as it mirrors the finding of higher scores for girls when using the Eneza platform, as reported by Kizilcec and Goldfarb (2019). Given that this finding has been observed in both studies, it would be valuable for future qualitative research to be carried out in order to understand in more depth the reasons behind these differences.

It is important to note, however, that quiz scores are not validated measures of learning outcomes, so should not be treated as such. Despite this, there is a technical value in considering differences in quiz scores, because of the role that this information plays within interactions with the system. Quiz scores are used by the system as a way of determining learners' progress through the content, so higher scores may indicate that girls in non-ASAL areas are moving through the content at a faster pace than girls in ASAL areas. This may also be indicated by the greater number of interactions logged by girls in ASAL areas. This suggests that this form of technology supports effective use for learners spanning a range of initial ability levels, by adjusting to learners' progress—so, for example, it could be used by learners in a mixed ability classroom.

There are some limitations to the present study and areas for further research. While the findings here suggest that the M-Shule platform is effective in promoting learning outcomes for both girls and boys, in areas of low and high connectivity (both ASAL and non-ASAL areas), the use of different scales for baseline and endline ASER data collection in the dataset used for secondary analysis prevented measurement of individuals' learning gains (comparisons could still be made across the treatment and control groups). One of the main potential advantages of using SMS as a medium for education is the prevalence of mobile phone subscriptions in Kenya, compared to levels of Internet access via computers (DataReportal, 2023). However, such statistics can mask a more complex picture in practice, with disparities between device ownership and subscription sharing. Furthermore, the presence of mobile devices does not mean that they will be accessible to children, and caregivers may have concerns about usage for educational purposes (Watson et al., 2023). Further research into the user perspective and practices would be useful.

There are limitations to both types of data drawn upon for the study, which includes server log data and remote literacy assessments. Server log data provide a comprehensive log of interactions with the system and generates a large dataset, but the level of insight which can be gained is restricted to the overall level of use, and logged quiz scores. Nonetheless, this provides an initial overview of use, and could be supplemented in the future with qualitative research to understand learners' perspectives and use of the content in practice. The reliability of data collection through telephone assessments has been demonstrated in similar contexts, particularly when collected for aggregate, impact evaluation purposes (Rodriguez-Segura & Schueler, 2022), however, it would also be useful to verify the findings through in-person assessments which would also allow further analysis at the individual level. The use of the ASER scale is also a potential limitation; while it has advantages and is frequently used as a non-resource-intensive assessment tool, the assumptions of the scale reflect controversies and debates in relation to the value of phonics and relationship with comprehension in the context of early years literacy (Bartlett et al., 2015).

The sample size is relatively small for a causal inference research design, and an evaluation at a larger scale would be beneficial and would allow the influence of mediating factors to be investigated with more certainty. The intervention here focused upon literacy, and it would be valuable to examine whether different subjects and topics are suitable for this medium. It is notable in this instance that the content adapted for this medium—the Tusome programme—had already been validated and proven to be effective. The initiative was undertaken as an emergency remote education measure during the Covid-19 pandemic; there are now open questions about how to use this mode of instruction alongside classroom provision. Given that the data here suggests that this form of EdTech can provide flexibility to a range of learners, leading to improved learning outcomes overall, the next steps need to address how it can be used to support learners in the return to school and at greater scale alongside classroom teaching.

ACKNOWLEDGEMENTS

This research was funded by the Foreign, Commonwealth and Development Office (FCDO) through the EdTech Hub (http://www.edtechhub.org). We would like to thank USAID for funding the Tusome implementation, and for permission to access the data for secondary analysis. We are grateful to Jasmin Baier at the Busara Centre for Behavioural Economics for preparing and sharing the data files for secondary analysis, and to Annette Zhao at Jigsaw Consult for initial analysis of the data. Many thanks to all of the caregivers and learners who have used the M-Shule platform and participated in the study. We are extremely grateful to Dr Ben Piper and Dr Alison Twiner for their feedback on an earlier version of this paper.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

In this study, we report findings from secondary analysis of data previously collected under the auspices of a different funder and programme, which are used here with permission. As such, we cannot publish the data for legal reasons.

ETHICS STATEMENT

Applications were made to the following bodies, and approval was granted: the EdTech Hub research ethics panel; the ethical review board of Maseno University, Kenya; and to NACOSTI for a research permit. The project was guided by the British Educational Research Association (BERA) ethical guidelines (BERA, 2018).

ORCID

Katy Jordan https://orcid.org/0000-0003-0910-0078

REFERENCES

Angrist, N., Aurino, E., Patrinos, H. A., Psacharopoulos, G., Vegas, E., Nordjo, R., & Wong, B. (2023). Improving learning in low- and lower-middle-income countries. *Journal of Benefit-Cost Analysis*, *14*, 55–80. https://doi.org/10.1017/bca.2023.26

Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020a). Stemming learning loss during the pandemic: A rapid randomized trial of a low-tech intervention in Botswana. Available at SSRN: https://doi.org/10.2139/ssrn.3663098

Angrist, N., Bergman, P., Evans, D. K., Hares, S., Jukes, M. C. H., & Letsomo, T. (2020b). Practical lessons for phone-based assessments of learning. *BMJ Global Health*, *5*(7), e003030. https://doi.org/10.1136/bmjgh-2020-003030

Angrist, N., Bergman, P., & Matsheng, M. (2020c). *School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic*. National Bureau of Economic Research. http://conference.nber.org/conf_papers/f145603.pdf

Bartlett, L., Dowd, A. J., & Jonason, C. (2015). Problematizing early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development*, *40*, 308–314. https://doi.org/10.1016/j.ijedudev.2014.10.002

BERA. (2018). *BERA ethical guidelines for educational research* (4th ed.). British Educational Research Association. https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018-online

Birch, I. (2018). *Economic growth in the arid and semi-arid lands of Kenya*. Institute of Development Studies. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14212

Chen, M., & Kizilcec, R. F. (2020). Return of the student: Predicting re-engagement in mobile learning. In *Proceedings of the 2020 Educational Data Mining Conference*. Online (virtual event). https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_95.pdf

Cotter Otieno, J., & Taddese, A. (2020). *EdTech in Kenya: A rapid scan*. EdTech Hub. https://doi.org/10.5281/ZENODO.3909977

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Crompton, H., Chigona, A., Jordan, K., & Myers, C. (2021). Inequalities in girls' learning opportunities via EdTech: Addressing the challenge of COVID-19. *EdTech Hub*. https://doi.org/10.5281/zenodo.4917252

DataReportal. (2023). *Digital 2023: Kenya*. https://datareportal.com/reports/digital-2023-kenya

Evans, D., & Popova, A. (2015). *What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews*. The World Bank. https://ideas.repec.org/p/wbk/wbrwps/7203.html

Evans, D. K., & Yuan, F. (2019). *Equivalent years of schooling: A metric to communicate learning gains in concrete terms (Working Paper No. 8752; World Bank Policy Research)*. World Bank. https://doi.org/10.1596/1813-9450-8752

Field, A. (2009). Chapter 15: Non-parametric tests. In A. Field (Ed.), *Discovering statistics using SPSS* (2nd ed., pp. 539–583). Sage.

Fonseca, J., Bahrawar, L., Dubeck, M. M., Sitabkhan, Y., Cummiskey, C., & Unadkat, D. (2023). *Girls have academic advantages and so do boys: A multicountry analysis of gender differences in early grade reading and mathematics outcomes*. RTI International. https://doi.org/10.3768/rtipress.2023.rr.0049.2305

Global Education Evidence Advisory Panel. (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. https://thedocs.worldbank.org/en/doc/231d98251cf326922518be0cbe306fdc-0200022023/related/GEEAP-Report-Smart-Buys-2023-final.pdf

Gupta, G., Sood, S., Roy, S., & Mehta, A. (2023). *Bharat survey for EdTech (BaSE) report*. Central Square Foundation. https://www.edtech.centralsquarefoundation.org/wp-content/uploads/BaSE-Report-web-version.pdf

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distance Learning*, *16*(3), 341–358.

Jordan, K. (2023). How can messaging apps, WhatsApp and SMS be used to support learning? A scoping review. *Technology, Pedagogy and Education*, *32*(3), 275–288. https://doi.org/10.1080/1475939X.2023.2201590

Jordan, K., Damani, K., Myers, C., & Zhao, A. (2023). The use of SMS and other mobile phone-based messaging to support education at scale: A synthesis of recent evidence. *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23)*, July 20–22, 2023, Copenhagen, Denmark. https://doi.org/10.1145/3573051.3596172

Jordan, K., & Myers, C. (2022). EdTech and girls education in low- and middle-income countries: Which intervention types have the greatest impact on learning outcomes for girls? *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22)*, June 1–3, 2022, New York City, NY, USA. https://doi.org/10.1145/3491140.3528305

Kizilcec, R. F., & Chen, M. (2020). Student engagement in mobile learning via text message. In *Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S 2020)* (pp. 157–166). Online (virtual event). https://doi.org/10.1145/3386527.3405921

Kizilcec, R. F., Chen, M., Jasińska, K. K., Madaio, M., & Ogan, A. (2021). Mobile learning during school disruptions in sub-Saharan Africa. *AERA Open*, *7*, 23328584211014860. https://doi.org/10.1177/23328584211014860

Kizilcec, R. F., & Goldfarb, D. (2019). Growth mindset predicts student achievement and behavior in mobile learning. In *Proceedings of the 6th ACM Conference on Learning @ Scale (L@S 2019)* (pp. 1–10). Chicago. https://doi.org/10.1145/3330430.3333632

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170–179). Leuven, ACM.

Laterite. (2021). *Research approval processes: Kenya, Tanzania, Ghana, Sierra Leone, Pakistan and Bangladesh. Report commissioned for the EdTech Hub*. EdTech Hub.

Leech, N. L., & Onwuegbuzie, A. J. (2002). *A call for greater use of nonparametric statistics*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Chattanooga, TN, November 6–8, 2002). https://files.eric.ed.gov/fulltext/ED471346.pdf

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

Mukoya, T. (2021). *Parents worry as crowded Kenyan schools reopen after coronavirus shutdown*. Reuters. https://www.reuters.com/article/us-health-coronavirus-schools-kenya-idUSKBN29912E

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Ngware, M., & Ochieng, V. (2020). *EdTech and the COVID-19 response: A case study of Kenya [Case Study]*. EdTech Hub. https://doi.org/10.5281/zenodo.4706019

Nicolai, S., Jordan, K., Adam, T., Kaye, T., & Myers, C. (2023). Toward a holistic approach to EdTech effectiveness: Lessons from COVID-19 research in Bangladesh, Ghana, Kenya, Pakistan, and Sierra Leone. *International Journal of Educational Development*, *102*, Article 102841. https://doi.org/10.1016/j.ijedudev.2023.102841

Piper, B., Destefano, J., Kinyanjui, E. M., & Ong'ele, S. (2018). 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

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), e12864.

Roberts, N., Mostert, I., & Plaatjies, L.-A. (2021). Measuring knowledge gains in an SMS m-learning intervention: The case of ChildConnect South Africa. In M. E. Auer & T. Tsiatsos (Eds.), *Internet of things, infrastructures and mobile applications* (pp. 69–80). Springer International Publishing. https://doi.org/10.1007/978-3-030-49932-7_7

Rodriguez-Segura, D., & Schueler, B. E. (2022). Can learning be measured by phone? Evidence from Kenya. *Economics of Education Review*, *90*, 102309. https://doi.org/10.1016/j.econedurev.2022.102309

Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448–467.

Sharples, M., & Pea, R. D. (2014). Mobile learning. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed.). Cambridge University Press.

Tembey, L., Baier, J., Ogolla, C., & Mohan, P. (2021). *Understanding barriers to girls' access and use of EdTech in Kenya during Covid-19 [Working Paper]*. EdTech Hub. https://doi.org/10.53832/edtechhub.0048

UNESCO Institute for Lifelong Learning. (2022). *M-Shule SMS learning & training, Kenya*. UNESCO Institute for Lifelong Learning. https://uil.unesco.org/case-study/effective-practices-database-litbase-0/m-shule-sms-learning-training-kenya

UNICEF Kenya. (2018). *Situation analysis of children and women in Kenya, 2017*. UNICEF. https://www.unicef.org/kenya/reports/situation-analysis-children-and-women-kenya-2017

Vagh, S. B. (2009). *Evaluating the reliability and validity of the ASER testing tools*. ASER Centre. http://www.linguaakshara.org/yahoo_site_admin/assets/docs/ASER-Reliability__Validity_Evaluation.11091338.pdf

Van Schoors, R., Elen, J., Raes, A., & Depaepe, F. (2021). An overview of 25 years of research on digital personalised learning in primary and secondary education: A systematic review of conceptual and methodological trends. *British Journal of Educational Technology*, *52*(5), 1798–1822. https://doi.org/10.1111/bjet.13148

Wagner, D. A. (2014). *Mobiles for reading: A landscape research review*. Working Papers (Literacy.org). https://repository.upenn.edu/literacyorg_workingpapers/10
Watson, J., Baier, J., Mughogho, W., & Millrine, M. (2023). An exploratory investigation into the factors related to EdTech use among Kenyan girls. *British Journal of Educational Technology*, *54*(4), 1006–1024. https://doi.org/10.1111/bjet.13307
White, H., & Sabarwal, S. (2014). *Quasi-experimental design and methods: Methodological briefs—Impact evaluation no. 8*. Innocenti UNICEF Office of Research. https://www.unicef-irc.org/publications/753-quasi-experimental-design-and-methods-methodological-briefs-impact-evaluation-no.html
Wolf, S., Aurino, E., Avornyo, E., Brown, A., & Tsinigo, E. (2021). Nudges to improve learning and gender parity: Supporting parental engagement and Ghana's educational response to Covid-19 using mobile phones. Presentation at the Society for Research on Educational Effectiveness (SREE) conference, Arlington, Virginia, USA, September 26–29, 2021.
Zubairi, A., Kreimeia, A., Jefferies, K., & Nicolai, S. (2021). *EdTech to reach the most marginalised: A call to action*. EdTech Hub. https://doi.org/10.53832/edtechhub.0045

How to cite this article: Jordan, K., Myers, C., Damani, K., Khagame, P., Mumbi, A., & Njuguna, L. (2024). Supporting equitable access to learning via SMS in Kenya: Impact on engagement and learning outcomes. *British Journal of Educational Technology*, *00*, 1–23. https://doi.org/10.1111/bjet.13533

APPENDIX

TABLE A1 Topics, subtopics, and number of unique learners per subtopic in Class 2 English and Kiswahili content.

Topic	Subtopic	Unique learners
Class 2 booster, English		
Welcome and greetings	Reading: Don at school	248
	Vocabulary: Welcome and greetings	91
	Grammar: Verb ‘to be’—Use of ‘am’ and ‘are’	61
Home	Reading: A new classroom for Joan	45
	Vocabulary: Home	33
	Grammar: Adding -s to show more than one	26
Myself (parts of my body)	Reading: It is good to be home	22
	Vocabulary: Parts of My Body	17
	Grammar: Use of ‘this’ and ‘that’	15
Weather and our environment	Reading: Mata and I	14
Class 2 booster, Kiswahili		
Karibu darasani	Kusoma Hadithi: Sikukuu ya Pasaka	140
	Msamiati: Maamkuzi	59
	Sarufi: Karibu darasani	40
Mimi na wenzangu	Kusoma Hadithi: Somo nilipendalo	29
	Msamiati: Mimi na wenzangu	21
	Sarufi: Mimi na wenzangu	18
Tarakimu 1	Kusoma hadithi: Safari ya Amina	14
	Msamiati: Tarakimu	11
	Sarufi: Kuambatanisha jina na nambari (Hadithi ya mwalimu)	10
Siku za wiki 1	Kusoma hadithi: Wimbo niupendao	8

TABLE A2 Topics, subtopics, and number of unique learners per subtopic in Class 3 English and Kiswahili content.

Topic	Subtopic	Unique learners
Class 3 booster, English		
School	Reading: Hare and Hyena	179
	Vocabulary: School	90
	Grammar: Use of the words 'was', 'were'	62
Activities in the home	Reading: Tami at the shop	47
	Vocabulary: Activities in the Home	41
	Grammar: Subject-verb agreement; was, were	34
Transport	Reading: The train trip	30
	Vocabulary: Transport	30
	Grammar: Use of the words 'me', 'you' and 'us'	27
Time and months of the year	Reading: A walk in the forest	25
Class 3 booster, Kiswahili		
Shuleni	Kusoma Hadithi: Wakati wa Likizo	119
	Msamiati: Shuleni	68
	Sarufi: Matumizi ya -ako, -enu	46
Haki zangu	Kusoma Hadithi: Shangwe Uwanjani	35
	Msamiati: Haki zangu 1	27
	Sarufi: Matumizi ya vizuri, vibaya	24
Usafiri	Kusoma Hadithi: Lishe Bora	19
	Msamiati: Usafiri	16
	Sarufi: Matumizi ya Herufi Kubwa	15
Familia	Kusoma Hadithi: Mzee Kichwa aita Mkutano	14

TABLE A3 Topics, subtopics, and number of unique learners per subtopic in Class 4 English and Kiswahili content.

Topic	Subtopic	Unique learners
Class 4 booster, English		
Activities at home and at school	Vocabulary: Activities at Home and at School	273
	Grammar: Doing words	108
Sharing duties and responsibilities	Vocabulary: Sharing duties and responsibilities	73
	Grammar: Somebody, nobody, anybody, everybody	58
Play time and sports	Vocabulary: Play Time and Sports	49
	Grammar: Words ending with -er and -est	41
Disease and food we eat	Vocabulary: Diseases and Food we Eat	34
	Grammar: Use of but, and, because	28
Technology	Vocabulary: Technology (Using a computer)	25
	Grammar: Use of will, shall	19
Class 4 booster, Kiswahili		
Marejeleo	Msamiati: Wanyama wapenda kazi	162
	Sarufi: Kinyume cha vitenzi	80
Shambani	Msamiati: Shambani	47
	Sarufi: Nafsi ya tatu—Umoja na Wingi	35
Sokoni	Msamiati: Sokoni	27
	Sarufi: Vihusishi ‘Nje ya’ na ‘Ndani ya’	23
Uzalendo	Msamiati: Uzalendo	20
	Sarufi: Vimilikishi -ake -ao na—etu	18
Miezi ya mwaka	Msamiati: Miezi ya Mwaka	17
	Sarufi: Matumizi ya kikomo(.)	16

TABLE A4 Results of regression analysis applied to the baseline data.

Predictor	Estimate	*SE*	Wald	*p*	Exp(*b*)	95% CI
Coefficients						
Grade	0.852	0.161	27.890	<0.001	2.344	1.708–3.215
Gender	−0.042	0.274	0.23	0.879	0.959	0.560–1.642
Location	0.154	0.276	0.310	0.578	1.166	0.678–2.005
Treatment	0.355	0.273	1.683	0.194	1.426	0.834–2.437
Intercepts						
1. Letter	−0.328	0.581	0.318	0.573	0.721	
2. Word	1.276	0.536	5.672	0.017	3.583	
3. Paragraph	2.857	0.571	25.065	<0.001	17.404	
4. Story	3.885	0.602	41.587	<0.001	48.650	

Note: No significant effect of treatment group at this stage.