Predictors of foundational learning for out-of-school children in Tanzania with interactive apps deployed direct to communities

Bethany Huntington

Submitted to the University of Nottingham

for the degree of Doctor of Philosophy

April 2024

Supervisors

Professor Nicola Pitchford, School of Psychology, University of Nottingham

Dr James Goulding, N/LAB, University of Nottingham

Abstract

The world is experiencing a global learning crisis, exacerbated by the COVID-19 pandemic, which has left over 393 children without basic literacy and numeracy skills. Educational technologies (EdTech) that promote autonomous learning may ameliorate this learning poverty. However, little is known about the mechanisms through which out-of-school children may acquire foundational skills using tablet-based technology and app-based learning software.

This thesis aimed to identify potential app, child, and community-level predictors of foundational learning outcomes following a large-scale EdTech intervention - the Global Learning XPRIZE (GLXP) competition - directly deployed to out-of-school communities in Tanzania. An exploratory, mixed-methods approach was adopted to inform a theoretical model of out-of-school app-based learning by identifying predictors at three different levels of explanation.

First, a comparative judgement experiment (Chapter 3) was undertaken to investigate app-level predictors, in which 41 non-expert participants compared five learning apps across 15 key pedagogical features. Results indicated that six pedagogical features were found to be most influential for learning—autonomous learning, motor skills, task structure, engagement, language demand and personalisation. EdTech has been shown to facilitate learning for girls, an at-risk population in low-income countries, so further research was needed to determine what app features may be most influential across genders. A gender and domain app study (Chapter 4) used machine learning methods and inferential statistics to identify which app features most predict girls' and boys' literacy and numeracy learning. Some app features, such as retrieval-based learning and engagement, were found to be broadly influential for learning. Five app features - engagement, autonomous learning, language demand,

personalisation, and curriculum links- showed a differential influence across genders and domains.

A machine learning regression approach was adopted to explore contextual predictors (Chapter 5) of literacy and numeracy learning. Child and community-level features were leveraged using competition survey and assessment data, and contextual covariates derived from open-source geospatial data. Reading habits, small family size, and high social connectedness were shown to be the most predictive factors for improved learning outcomes. Community factors were found to be more predictive of learning improvements than child-level features, highlighting the importance of infrastructure.

Results from the GLXP imply that EdTech deployed to out-of-school children living in remote villages can support autonomous learning. However, no qualitative data regarding implementation was gathered during the competition to draw firm conclusions. Therefore, an expert elicitation (Chapter 6) explored the broader implementational impact and challenges experienced, whereby 14 key informants were interviewed about the autonomous learning process. Four key themes were generated: ‘Technology as a novel concept’, ‘Children don't learn in a vacuum’, ‘Respecting the cultural context’ and ‘Accessibility problems in a mobile world’. Community support was prominent throughout the competition, emphasising its importance for learning with EdTech and raising questions about whether out-of-school children can learn autonomously with this technology.

Key implications and recommendations are outlined for technology developers, educators, and policymakers to consider when designing and implementing app-based learning interventions for out-of-school children. If influential mechanisms are carefully incorporated into the design and implementation of EdTech interventions, they may support even the most marginalised learners.

Executive summary

Background

Learning foundational literacy and numeracy skills is crucial to living a healthy and productive life. However, a learning crisis is being experienced worldwide, creating significant inequalities in access to good-quality schooling, with girls among those most at risk. Across Sub-Saharan Africa alone, over 98 million children are illiterate, innumerate, and do not attend school. Educational technologies (EdTech) that promote autonomous learning are an alternative solution to traditional schooling methods that may help to address this learning poverty.

Purpose

Little is known about if and how these technologies can be effectively implemented within communities to support out-of-school children in learning foundational literacy and numeracy skills. To address this knowledge gap, this thesis used a unique large-scale EdTech intervention, the Global Learning XPRIZE competition, as a case study. The purpose of this thesis was to explore the mechanisms through which out-of-school children learn when using interactive EdTech. The primary aim was to establish which app-level, child-level, and community-level features are most predictive of improvements in foundational learning with an EdTech intervention.

Methods

This thesis adopted a mixed-methods approach to explore predictors at three different levels of learning: app, child, and community. The XPRIZE Foundation provided pre-and post-intervention datasets, which reported on Early Grade Reading Assessment (EGRA) and Early Grade Maths Assessment (EGMA) to assess learning over time. Data from a contextual survey was also provided, which provided information about the home environment, family life, and previous experiences with technology and schooling of the children and caregivers participating in the study.

Four studies were conducted that report findings of predictive factors at the app level (studies 1 and 2), child level (study 3), and community level (studies 3 and 4). Gender and domain differences were also explored at the app level to identify which app features most predict girls' and boys' literacy and numeracy learning and determine whether there are significant gender differences in app-based learning for out-of-school children (study 2). To explore these predictors at each level, an experimental comparative judgement task, secondary data analyses using machine learning methods, open-source geospatial data and inferential statistics, and a qualitative expert elicitation were conducted. The results were used to inform a holistic multi-level model of EdTech learning for out-of-school children.

Key findings

App-level predictors

- Six pedagogical features were shown to predict foundational learning: autonomous learning, motor skills, task structure, engagement, language demand, and personalisation.
- Of the six app features, motor skills and task structure were found to be broadly influential for learning. The other four app features - engagement, personalisation, language demand and autonomous learning - showed significant differences across genders and domains. An additional feature, curriculum links, also showed significant differences, although this app feature was less influential overall.
- All three accessibility-based app features (autonomous learning, language demand and motor skills) were found to be highly influential in learning improvements. This could be because the demographic is likely to have high proportions of children with undiagnosed SEND, identified by a large number of ‘low-achievers’ in the sample.

- Gamification and free play were not found to significantly influence foundational learning.

Child-level predictors

- At the child level, reading habits, prior school attendance, and coming from a small family were the most predictive factors for improved learning outcomes.

Contextual predictors

- Community-level factors were found to be substantially more important than child-level factors for predicting learning improvements. This highlights the importance of infrastructure and implementation in an EdTech intervention.
- The most influential community-level factors were those centred around social connectedness and physical isolation, such as proximity to health centres and police stations.
- Considerable community support was provided to children when using EdTech.
- Respecting the cultural context in which the intervention is implemented was found to be pivotal. As the technology was novel, there were also initial barriers faced due to wariness among community members.

Policy Implications and Recommendations

1. **App design—**Broadly influential app features, such as engagement and retrieval-based learning, should be incorporated in all app designs for out-of-school children. App designers should prioritise these features for their broad applicability, while also considering cultural adaptability. Rather than using non-generic features narrowly targeted at specific groups, developers should ensure that apps are inclusive by design, offering accessibility features to prevent barriers for children with SEND and diverse learners.

2. **Infrastructure—**While some communities may be too remote to benefit fully from EdTech interventions due to a lack of infrastructure, the types of infrastructure that influence learning vary widely and merit further exploration. Interventions should focus on leveraging existing infrastructure where possible to optimise cost-efficiency.
3. **Implementation—**EdTech solutions should leverage community engagement to facilitate out-of-school children's learning with EdTech, recognising that community support may play a critical role. Sensitisation processes that involve local educators and community leaders can help ensure the intervention is contextually relevant and culturally accepted. Flexible implementation strategies that adapt to each community's needs and resources are recommended to foster sustainable support and reduce wariness towards new technology.

Further research is needed to determine if these influential predictors of foundational learning with out-of-school children using interactive EdTech can be generalised to different contexts. Evaluating whether learning gains can be sustained over time is pivotal, as this will help decision-makers assess whether investment in EdTech is a cost-effective and efficient method to facilitate foundational learning acquisition for out-of-school children.

Publications

Work from this thesis has been published in the following articles:

Chapter 3:

Huntington, B., Goulding, J., & Pitchford, N. J. (2023). Pedagogical features of interactive apps for effective learning of foundational skills. *British Journal of Educational Technology*, *54*(5), 1273–1291. https://doi.org/10.1111/bjet.13317

Chapter 6:

Huntington, B., Goulding, J., & Pitchford, N. J. (2023a). Expert perspectives on how educational technology may support autonomous learning for remote out-of-school children in low-income contexts. *International Journal of Educational Research Open*, *5*, 100263. https://doi.org/10.1016/j.ijedro.2023.100263

Conference Presentations

Huntington, B., Goulding, J., & **Pitchford, N.J**. (2023). The influence of pedagogical features and implementation on the success of mobile technologies for foundational math interventions with out-of-school children. *mEducation Alliance Symposium 2023.* Washington DC, USA.

Huntington, B., Goulding, J., & Pitchford, N.J. (2023). A mixed methodological approach to determine the efficacy of digital technologies in educating marginalised, out-of-school children. *The Midlands Graduate School ESRC DTP Annual Conference.* Birmingham, UK.

Huntington, B., Goulding, J., & Pitchford, N.J. (2023). Pedagogical features of interactive apps for effective learning: A comparative judgement study with out-of-school children [invited virtual talk]. *SIG 15 Brown Bag Seminar.* Online series.

Huntington, B., Goulding, J., & Pitchford, N.J. (2022). The importance of partnerships for successful deployment of EdTech interventions in remote, low-income contexts: Lessons from the GLXP. *BAICE Conference*. Edinburgh, UK.

Huntington, B., & Pitchford, N.J. (2022). Cross-cultural insights into how tablet technology can support children with special educational needs and disabilities to acquire foundational skills: A mini-review. *JURE Conference.* Porto, Portugal.

Huntington, B., Goulding, J., & Pitchford, N.J. (2021). Transforming Global Learning with Digital Technologies: A qualitative exploration of the use of educational technology with marginalised, out-of-school children living in remote settings [virtual]. *The Midlands Graduate School ESRC DTP Annual Conference.* Birmingham, UK.

Huntington, B., Goulding, J., & Pitchford, N.J. (2021). Transforming Global Learning with Digital Technologies [virtual poster session]. *University of Nottingham Psychology PGR Conference.* Nottingham, UK.

Huntington, B., Goulding, J., & Pitchford, N.J. (2021). Transforming Global Learning with Digital Technologies: A qualitative exploration of the use of educational technology with marginalised, out-of-school children living in remote settings [virtual]. *EDULEARN21 Conference.* Palma de Mallorca, Spain.

Funding

This thesis was funded by the Economic and Social Research Council [grant number ES/J500100/1] and supported by the XPRIZE Foundation.

Dedication

This thesis is dedicated to my late Mommar, Dianne Lowe, with whom I started this adventure, but she never got to see me finish. Losing you made this PhD feel 1000 times harder, but your unwavering belief that I could do it spurred me to the finish line. I hope I have made you proud. I love you.

Acknowledgements

Firstly, thanks to my supervisors, Professor Nicola Pitchford and Dr James Goulding, for your guidance, support, and encouragement throughout this PhD. Nicola, thank you for providing thorough feedback, valuable planning and results sessions, consistent pastoral support when I was finding things tough, and help planning my future career. I would not have finished without you convincing me to pull myself together on a somewhat regular basis! James, thank you for introducing me to a whole new area of research in data analytics and working with me until I felt confident—a big difference from the complete avoidance at the start! Thank you for the pep talks in our weekly meetings and your endless patience as you explained a machine-learning algorithm to me for the tenth time.

Thank you to everyone in the psychology department and N/LAB who has helped me through academic support, encouragement, or teaching/research opportunities. I will always appreciate how welcoming and helpful everyone was, even though I did not spend much time in the offices.

Thank you to the participants in my two experimental studies; this work would have been impossible without people giving up their time to share their perspectives with me. I am also thankful to the XPRIZE Foundation, particularly Dr Emily Church, for supporting and facilitating this research.

I am so thankful for my family and friends, who have kept me going for five years. Thank you to my soon-to-be in-laws, the McJenkPratts, for always checking in, supporting me, and providing a great source of entertainment through the lockdown era. Special thanks to Sam and Jordan, who offered emotional support and encouragement when it was very much needed.

Bec, thank you for being here throughout all the highs and lows of the last eleven years since we met on Freshers week of our undergraduate year. There is no one I would rather have by my side as we have laughed and cried our way through it all.

Thank you to my family - Mum, Sam, Ben, Keira, Lillie, Gran, and Grandad. Ben and Keira - you deserve a medal for living with me throughout the pandemic years of this PhD. Thank you for your support, patience, and encouragement during a highly stressful time. Gran and Grandad - you still do not have a clue what this PhD involves, but you have been relentlessly supportive throughout, and there has never been a doubt in your mind that I would manage it. Your unwavering faith in me has been a huge motivation throughout this. Mum - you have been my number one supporter my whole life, and I could not be more thankful. Thank you for listening to me, calming me down, and providing tough love when needed. Thank you for always believing in me; I hope I have made you proud.

A special thanks to my fiancé, Callum. This PhD has been a turbulent time for me, and I could not have gotten through it without you. You have been remarkably kind, patient, and encouraging. Thank you for the endless pep talks, the daily writing targets, and listening to me babble on about research you know nothing about. I am so grateful you have been alongside me for every step of this journey, celebrating every little win with me. The last five years have been the most stressful of my life but also the best due to everything we have accomplished together. Finally, I thank Mabel for being the best companion through long months of writing and working from home.

Table of Contents

Abstract **2**

Executive summary **4**

Publications **8**

Conference Presentations **8**

Funding **9**

Dedication **10**

Acknowledgements **10**

Chapter 1: Introduction and Literature Review 21

Background 21

1) The capability of out-of-school children to engage with and learn from tablet-based technology and app-based learning software 24

Educational challenges for out-of-school children 24

Existing EdTech initiatives for out-of-school children 26

Consideration of gender inequalities 44

Children with Special Educational Needs and Difficulties (SEND) 50

2) The mechanisms through which out-of-school children may acquire foundational skills using tablet-based technology and app-based learning software. 52

Challenges with implementation and infrastructure 58

Edtech deployment during the COVID-19 pandemic 61

Theoretical framework 66

Thesis Aims **72**

Chapter 2: Methodology **74**

Philosophical position of the research 74

Justification of methodological approach 75

Global Learning XPRIZE competition: Case Study 80

The learning apps 83

The main dataset 89

Overview 89

Literacy and numeracy data 90

Contextual data 91

Survey data descriptives 91

Summary 94

Chapter 3: Comparative Judgement of Pedagogical Features in Interactive Apps **100**

Background 100

Importance of designing effective learning apps 100

Effective pedagogical features 101

Current study 107

Method 109

Design 109

Participants 109

Apparatus and Materials 110

Procedure 112

Data Analysis......113

Results......115

1) Which features characterise each of the five finalist learning apps used in the GLXP? 115

2) Do the five finalist learning apps of the GLXP, that have been shown to support positive learning outcomes with out-of-school children within the same RCT, share features?......119

Discussion......120

Limitations......124

Conclusion......124

Chapter 4: Relative Importance of App Features across Domains and Gender 126

Background......126

App features and learning outcomes......126

Gender differences......127

Can girls learn with apps in low-income countries?......128

Gender differences in app features......128

Current study......130

Method......132

Dataset......133

Generating lambda scores - the Bradley-Terry model......133

Data cleaning for modelling......134

Modelling......137

Inferential analysis......141

Results 143
Bradley-Terry analysis 143
Classification models...146
Inferential tests 154
Discussion 163
Novelty of research methods 167
Limitations...168
Conclusion...169
Chapter 5: Contextual Predictors of Foundational Learning Outcomes 171
Background 171
Current Study 173
Child features...174
Village features (also referred to as community-level factors) 176
Child and village features combined 178
Method 179
Design..179
The data 180
Feature Engineering...181
Modelling Approach..188
Results 189
RQ1: Which factors best predict foundational learning skills in out-of-school children in remote areas of Tanzania? 189

RQ2: Which factors best predict improvements in foundational learning outcomes after implementing an educational technology intervention with out-of-school children in remote areas of Tanzania? 194

Discussion 199

Limitations 207

Implications and conclusions 208

Chapter 6: Expert Elicitation of the GLXP competition 210

Background 210

Autonomous learning for out-of-school children through EdTech 211

Current study 212

Method 213

Research Design 213

Ethics approval 214

Participant Recruitment 214

Participants 214

Data Collection Procedure 215

Data Transformation 215

Data analysis 216

Researcher characteristics and reflexivity 217

Analysis & Preliminary Discussion 217

Theme 1: Technology as a novel concept 217

Theme 2: Children don’t learn in a vacuum 222

Theme 3: Respecting the cultural context 228

Theme 4: Accessibility problems in a mobile world 229

Discussion 232

Technology as a novel concept 232

Children don't learn in a vacuum 234

Respecting the cultural context 236

Accessibility problems in a mobile world 236

Limitations 237

Conclusion and Future Directions 238

Chapter 7 - General Discussion 240

Level of explanation: App features 242

Level of explanation: Child (context) 244

Level of explanation: Community (context) 245

Infrastructure 245

Implementation 252

Levels of explanation not explored in this thesis 253

Level of explanation: Macro 254

Level of explanation: Chrono 256

Theoretical model of EdTech learning for out-of-school children 260

Value of mixed-methods approach 266

Conclusion 268

References **.. 270**

Appendices **.. 325**

List of tables

Table 1 Implementation data extracted from the five studies identified in the literature search for EdTech interventions with out-of-school children 33

Table 2 Brief description of the five finalist learning apps from the GLXP as described on the organisations' websites (retrieved October 2022) 84

Table 3 Summary of research questions that are addressed in this thesis 96

Table 4 App features that have been attributed to support children's learning in the literature 102

Table 5 Binomial test of probability results for each of the five finalist learning apps 116

Table 6 Spearman's rho correlation matrix between apps across the 15 features examined 120

Table 7 Lambda scores for each app feature 144

Table 8 Summary of the mean Random Forest performance for literacy and numeracy improvement per gender 146

Table 9 Mean (SD) SHAP values from the 30 Random Forest classification models and their overall rank in influencing the prediction of whether a child achieves foundational skills following an EdTech intervention 148

Table 10 Results of the 2x2 ANOVAs for each app feature: F-values, p-values, effect sizes and the corresponding post-hoc t-test comparisons for significant interactions 155

Table 11 A complete list of the child and village-level features generated and their data source 184

Table 12 Summary of the performance of the six regressor models conducted to determine factors that best predict foundational learning skills in out-of-school children in remote areas of Tanzania (RQ1). 190

Table 13 Summary of the performance of the six regressor models conducted to determine factors that best predict improvements in foundational learning skills after implementing an Edtech intervention with out-of-school children in remote areas of Tanzania (RQ2). 195

Table 14 Quotes to support Theme 1: Technology as a novel concept................................218

Table 15 Quotes to support Theme 2: Children don't learn in a vacuum 222

Table 16 Quotes to support Theme 3: Respecting the cultural context 228

Table 17 Quotes to support Theme 4: Accessibility problems in a mobile world 230

List of figures

Figure 1 Bronfenbrenner's Ecological Systems Theory 67

Figure 2 Outhwaite's multi-level model outlining factors that may influence learning outcomes following a maths app intervention 69

Figure 3 A holistic multi-level model depicting the potential predictors of learning improvement for out-of-school children using an EdTech intervention.................................71

Figure 4 Histogram showing the age distribution of participants at baseline 92

Figure 5 A histogram showing the raw EGRA improvement scores after the GLXP 135

Figure 6 A histogram showing the raw EGMA improvement scores after the GLXP 136

Figure 7 Box plots depicting the distribution of the 30 SHAP values for each gender per domain...152

Figure 8 Mapping the 12 predictor tasks to literacy and numeracy baseline and improvement scores for child and village features...180

Figure 9 SHAP summary plots for the six experiments conducted to determine factors that best predict basic literacy and basic numeracy skills in out-of-school children in remote areas of Tanzania (RQ1) 193

Figure 10 SHAP summary plots for the six experiments conducted to determine factors that best predict improvements in foundational learning skills after implementing an Edtech intervention with out-of-school children in remote areas of Tanzania (RQ2). 198

Figure 11 Thesis results embedded within the multi-level model for predictors of learning improvement for out-of-school children using an EdTech intervention................................264

Appendices

Appendix A: Search strategy used to identify out-of-school learning intervention studies for Chapter 1 - Introduction & Literature Review..325

Appendix B: Contextual survey questions that the XPRIZE Foundation asked the children and their caregivers at the pre-test and/or post-test sessions for the GLXP competition 326

Appendix C: List of variables reported in the primary XPRIZE dataset relating to the EGRA/EGMA tests...329

Chapter 1: Introduction and Literature Review

Background

Acquiring foundational literacy and numeracy is a fundamental human right, yet in 2021, it was estimated that over 393 million children of primary school age failed to gain basic literacy skills and perform simple mathematics (Save The Children, 2021; UNESCO, 2019). This learning crisis was exacerbated further by the global COVID-19 pandemic, as 100 million more children are now thought to be under the minimum proficiency level for reading, with an estimated 70% of 10-year-olds unable to understand a simple written text (UNESCO, 2021; UNICEF, 2022a). Even before the pandemic, learning poverty was high at 57%, with global progress stagnant since 2015 (World Bank et al., 2022). This indicates that many children do not possess the basic foundational skills required to live a healthy and productive life. This crisis urgently needs addressing as it perpetuates significant inequalities in access to and provision of quality education, costing governments upwards of $129 billion a year globally. Around 10% of global spending is currently thought to be wasted on 'poor education' which fails to produce the desired learning outcomes for children (UNESCO, 2019; World Bank, 2019).

While this learning crisis impacts countries across the globe, low-to-middle-income countries (LMICs) are disproportionately impacted, with a one-third increase in learning poverty. Sub-Saharan Africa is particularly affected, where poverty levels are high, traditional methods of education have been insufficient to address the problem, and 98 million children do not attend school (UNESCO, 2022c). Children living in remote villages, girls, and children with learning difficulties are among the most affected by the education crisis in low-income countries (Adukia & Evans, 2023; World Bank, 2022a; UNICEF, 2021a). They have lower school enrolment and completion rates leading to higher out-of-school numbers than other populations, highlighting the need to prioritise these intersecting groups. Educational challenges result in social and financial dependency, limit the extent to which individuals can actively participate in society, and raise vulnerability to pernicious social issues such as forced marriage, female genital mutilation (FGM) and child labour (International Centre for Research on Women, 2016). This consequently negatively impacts population health and well-being and a country's economic growth potential.

It has been argued that traditional education methods have failed to solve this crisis and that innovative, alternative approaches are required (World Bank, 2018). One such approach is mobile educational technology (EdTech) and interactive apps. There is an emerging evidence base supporting the potential of EdTech with interactive apps to provide access to high-quality education globally (Bettinger et al., 2020; Tauson & Stannard, 2018), with their use potentially guarding against learning loss in future pandemics (OECD et al., 2022).

Interactive apps can also promote autonomous learning, which is conceptualised as an individual's ability to take control of their learning by supporting the management of what they learn and when, engaging reflectively in the learning process and evaluating their own progress (Holec, 2001; Lan, 2018). Autonomous learning is a valuable skill when using EdTech in large classrooms (Jordan et al., 2021) and a perceived necessity when the learner is out of school and receives no formal support or instruction (Huntington et al., 2023b). Thus, children in LMICs receiving poor or no formal education might benefit from a learning app intervention to acquire core foundational skills. However, the effectiveness of learning technology remains under-researched and under-documented, particularly for out-of-school children in Sub-Saharan Africa. As a result, it has yet to be determined if and how EdTech might be deployed successfully directly to communities to promote the learning of foundational skills in out-of-school children.

To date, most EdTech research has employed experimental studies to generate an evidence base measuring the effectiveness of learning apps within different populations, situations, and learning styles, which has been used to inform practice (Spieth et al., 2016). A recent meta-analysis demonstrated that technological interventions can produce significant improvements in attainment and learning outcomes for foundational skills compared to standard practice in LMICs, along with enhanced attitudes and motivation through app personalisation (Major et al., 2021; Jones et al., 2013). Improvements in learning outcomes for foundational skills have been evidenced in different LMICs in the Global South (Lurvink & Pitchford, 2023) and across different groups of learners, such as low and high achievers (Bardack et al., 2023), speakers of more than one language (Outhwaite et al., 2020), and children with special educational needs and disabilities (Pitchford et al., 2018; Lurvink & Pitchford, 2023). In contrast, some studies report no effect of EdTech on language skills (e.g. Araya et al., 2019; Carrillo et al., 2011; Lai et al., 2013) and motivation (e.g. Ito et al., 2021), and suggest that EdTech can increase maths anxiety (Araya et al., 2019), which questions the efficacy of EdTech in raising learning outcomes. Furthermore, most EdTech research has been conducted within a school environment where teachers are available to scaffold and support children's learning as needed (e.g. Lurvink & Pitchford, 2023). A review of EdTech research suggests that self-led or autonomous learning interventions are the most effective at raising learning outcomes, even within a school environment, which indicates that EdTech using self-led interventions may be highly valuable for out-of-school children with no formal educational support to learn foundational skills (Rodriguez-Segura, 2022).

Accordingly, research is needed to inform an evidence-based understanding of 1) the capability of out-of-school children to engage with and learn from tablet-based technology and app-based learning software and 2) the mechanisms through which out-of-school children may acquire foundational skills using these methods, identifying key factors that may predict successful learning outcomes. These insights are critical to guiding educational app development for this population and informing educational policies that address this global learning crisis, which is becoming increasingly significant due to rising out-of-school numbers and the learning losses experienced through the pandemic. This thesis addresses these two research gaps by first establishing what is known about out-of-school children and how they learn (current chapter) before employing mixed methods to further investigate the mechanisms through which marginalised, out-of-school children may use tablet-based technology to acquire foundational skills (Chapters 3-6).

1) The capability of out-of-school children to engage with and learn from tablet-based technology and app-based learning software

To understand the capability of out-of-school children to learn with tablets and app-based software, it is first pivotal to consider the unique challenges that this population may face and the potential impacts this may have on learning interventions. Existing initiatives that have been conducted with learning apps and this population will also be examined. As it has been established that girls in rural communities are particularly high risk for being out-of-school and having lower literacy levels, the role of gender will be explored, concentrating on gender discrepancies experienced in this context, gender differences found when using educational software as a learning tool and the importance of increasing educational access for girls. Finally, existing research for children with SEND is explored, as another high-risk group that is likely to be impacted by inequalities in school access.

Educational challenges for out-of-school children

Out-of-school children living in marginalised communities are likely to face a series of unique challenges not experienced by those who usually attend formal schooling, some of which may impact their ability to learn using tablet-based technologies. One of these challenges is that many out-of-school children live in rural, low-income areas, so may have limited or no access to technological devices like smartphones or tablets to support their education (Kukulska-Hulme, 2023; Rodriguez-Segura, 2022). When they do have access, the devices may belong to the household unit, limiting their independent use and potential to be used as an educational tool (Duby et al., 2022; Krönke, 2020; Yardi & Bruckman, 2012). Such low penetration of technology could limit users' digital literacy and familiarity with the hardware on which EdTech interventions are often deployed, which may have a significant impact on how an intervention is received and, ultimately, its effectiveness (Rodriguez-Segura, 2022; Garcia, 2019).

Another challenge faced by out-of-school children in some low-income countries is the lack of access to electricity, with 53% of Sub-Saharan Africa and 600 million people living without any access to electricity, and millions more anticipated to live with unreliable, limited electricity access (United Nations, 2023). Without electricity, children would struggle to follow an EdTech programme due to a lack of charging facilities. However, EdTech initiatives focused on low-income and out-of-school environments are beginning to address these two challenges as part of their learning provision, as some prominent learning apps are fully functional without grid electricity. For example, a custom device named the 'onetab' has been created by onebillion, a non-profit organisation that builds scalable educational software, and this comes with optional solar-powered chargers (onebillion, 2022). While the organisation advertises the additional charger option as low-cost, funding is already a critical concern when considering EdTech programmes, particularly in low-income countries where governments are already lacking the financial resources to meet educational needs (onebillion, 2022; UNESCO, 2023). As they need to prioritise funding effectiveness, buying chargers for each individual may be too costly in addition to the tablet-based technology itself. An alternative option was introduced by the GLXP competition, an initiative testing

EdTech software solutions with out-of-school children in remote areas of Tanzania. They installed local charging stations for small groups of villages to access as and when needed (XPRIZE, 2019). Prominent infrastructural challenges remain, but these examples indicate that developers are slowly beginning to offer innovative solutions to address key problems as part of their technological provision.

Another crucial difference between out-of-school children and those in formal schooling or temporarily learning from home is the level and quality of support they receive alongside their learning. Children who were out of school because of COVID-19 restrictions still received regular support from their class teachers, usually through online teaching sessions (Crompton et al., 2021a; Goudeau et al., 2021; Carroll & Constantinou, 2023). Most EdTech interventions implemented in low-income countries include at least in-person support elements, whether from teachers, volunteers, parents or group learning centres and sessions (Pitchford, 2015; Kaye et al., 2020; Hennessy et al., 2021). While out-of-school children may receive informal support from their parents and other family members, it is essential to note that in rural, low-income settings, caregivers are also likely to have low literacy levels and, therefore, be unable to provide the complex, evaluative feedback needed to support learning (Fute et al., 2022; Cayton-Hodges et al., 2015). Previous research has suggested that EdTech may only be successful for children if combined with comprehensive in-person pedagogy and supported scaffolding, so this lack of educated support may substantially impact learning outcomes with out-of-school children (Rodriguez-Segura, 2022; Habyarimana & Sabarwal, 2018).

Existing EdTech initiatives for out-of-school children

Several initiatives have already been undertaken in LMICs that use EdTech solutions to tackle the global learning crisis with out-of-school children, with mixed levels of success due to the unique and extensive challenges this demographic faces. Low-technology options like radio and television have successfully supported foundational learning outcomes for rural, out-of-school children when carefully designed with the specific context in mind (Thinley & Rui, 2023). Compared to other technologies, such as smartphones and tablets, low-technology options are considered a cost-effective option with high initial costs but lower recurring costs, which are minimised further at scale (Zacharia, 2020).

Established interactive radio instruction (IRI) initiatives are commonly utilised in schools, providing educators with available, high-quality content where class sizes are large and under-resourced. An example is the Rising on Air (ROA) programme, which currently serves 250,000 students across more than 700 schools in Sierra Leone, Liberia, Ghana and Rwanda (Rising Academies, 2023). This 20-week structured curriculum content delivered by radio scripts alongside teacher training support has shown success in 25 countries as a resource to teach content such as foundational literacy, numeracy and language to young children and adolescents (Afoakwah et al., 2021; Rising Academies, 2020). A 12-week randomised control trial utilising the ROA programme and focusing on foundational maths skills was implemented in 15 Rising Academies low-cost private primary schools in Ghana with 1,359 children aged 8-11 (719 who were in the intervention group; Afoakwah et al., 2021). No significant differences were found in numeracy skills between those who participated in the intervention and those who did not, suggesting that the intervention was not significantly beneficial for improving numeracy skills. Despite the intervention group having a balanced gender split (352 girls vs 367 boys), gender effects were not explored for learning improvements. Students reported that they enjoyed the sessions, but engagement was also a significant challenge for both genders, even when the planned lessons were halved in length to address the low engagement (Afoakwah et al., 2021). As demonstrated, the ROA programme has had mixed success in schools in low-income settings, and the RCT in Ghana provided key insights to suggest that it may be difficult to effectively engage children in radio-based learning provision. It has been suggested that to maximise chances of success, low-tech solutions should encourage more active student interaction throughout (Afoakwah et al., 2021), which would be even more challenging in an out-of-school setting without trained teachers and additional resources.

Another prominent example of out-of-school IRI is the USAID-supported Taonga Market programme created to support Zambia's rising out-of-school population with high-quality education and psychosocial support (Education Development Center, 2020). This support was based on custom-made radio learning centres staffed by volunteers from the local communities (which may be a physical structure or convenient location to gather, such as a tree or local landmark; Williams, 2022). The programme consisted of 150 lessons for each grade level, each consisting of a 30-minute broadcast and accompanying worksheet activities (Sitali, 2006). The initiative was highly successful, reaching over 62,000 out-of-school orphans and vulnerable children aged 5-19 years in the initial pilot (USAID, 2009). The Zambian Ministry then rolled the programme out to 3000 community learning centres, leading to 1.2 million students outperforming their peers in formal schools in reading proficiency (UNESCO, 2024a). Results showed children receiving the IRI intervention to score 14.3% higher in literacy than those without this provision, and out-of-school children were found to perform at least slightly better than control groups across all subjects tested (literacy, mathematics, science, and social studies) after receiving IRI (Education Development Center, 2020; USAID, 2009). As with the Rising on Air programme, the learner population was almost equally divided by gender (51.1% girls vs 49.9% boys), yet gender effects were not explored. This initiative is highly beneficial in demonstrating that children without formal schooling can improve their foundational learning skills, but the children involved were based in learning centres with adult support and supervision from a dedicated mentor (Williams, 2022). Therefore, the learning is not autonomous, and it remains unclear whether this IRI initiative would be as successful with children who have to work completely autonomously within their communities without the informal support offered in the Taonga Market programme.

While many low-technology options, such as the IRI solutions explored, are a popular choice for LMICs to address inequalities in out-of-school access to education, it is pivotal to note that less than 40% of LMIC households have personal access to a radio or television (Mundy & Hares, 2020), suggesting that it is not a sustainable, long-term solution if the hardware is not provided. While this has been somewhat addressed by existing programmes creating informal school settings to broadcast the sessions (as in the Taonga Market programme; Education Development Center, 2020), large groups of children in one setting may be too distracting for the children trying to learn, making it extremely difficult for them to concentrate and hear the material. Furthermore, a disadvantage of radio and television-based learning is that the content and progression rate are not personalised to the individual learner, which is an approach shown to positively influence the impact of technology-based learning (Major et al., 2021).

Furthermore, each session is broadcast at a scheduled time, which the learners need to be available for and cannot miss to ensure the continuity of learning content (Williams, 2022). Having fixed broadcast times provides a prominent barrier to access for some children, particularly girls from low-income countries who typically have higher familial responsibilities and domestic workloads within their households (Njie et al., 2015; Hossain et al., 2023). Despite the IRI solutions presented having the opportunity to explore gender effects, they did not evaluate potential differences, which is problematic for trying to establish whether the radio-based sessions are broadly successful at improving learning outcomes or more suitable for one gender than another due to barriers for access. Gender-based equality in access to education is already a prominent concern for low-income countries (Crompton et al., 2021c), and providing scheduled radio provision as an out-of-school solution to learning could further exacerbate the evidenced inequalities. To ensure that girls are sufficiently targeted alongside boys, EdTech learning solutions should ensure that any scheduled activities are at suitable times to enable participation for all genders (Samuels et al., 2022). Alternatively, utilising a higher technology solution, such as tablet-based technology, could provide a more comprehensive, inclusive, and high-quality solution due to the high flexibility and ability to personalise instruction for the recipients.

One popular programme, One Laptop Per Child (OLPC), focused on higher technology hardware provision. Governments and NGOs aimed to increase access to education by simply providing children in LMICs with their own laptops, either directly to the student (as supplementary out-of-school provision) or providing classrooms with enough hardware that each child had access to one (Rodriguez-Segura, 2022). Autonomous learning with the technology was heavily promoted throughout the programme. The first large-scale RCT evaluating OLPC was conducted in 319 schools with 4,100 primary-aged students in rural Peru over 15 months, in which literacy, numeracy, language, and cognitive skills and time spent on technology were measured (Cristia et al., 2012). It was found that children in the experimental condition showed significantly higher improvement than controls for fluid intelligence and overall cognitive skills. In contrast, no significant improvement was found in mathematics and language skills, as well as verbal fluency, decoding and overall academic achievement. Taking part in the programme also led to a notable increase in the use of technology outside of learning hours (Cristia et al., 2012).

The OLPC initiative has been heavily criticised due to highly mixed results across the programme - for example, significant improvement was shown for maths and not reading when the programme was tested in Uruguay (de Melo et al., 2014) - and the lack of evidence of improvements in the core subjects of maths and literacy (Rodriguez-Segura, 2022;

Paradowski, 2015). There is evidence of some improvement as a result of OLPC, which has been highlighted, but this was using a 10% significance level, which is lenient and thus heightens the risk of Type 1 errors (false positives) and claiming statistical significance for marginal differences (Park, 2003; Cristia et al., 2012; de Melo et al., 2014). There were also no effect sizes reported, which are valuable measurements to provide comparative benchmarks for similar research studies and to establish the practical significance of findings (Funder & Ozer, 2019), which is crucial when trying to establish whether an EdTech solution could be implemented across different populations and settings.

While the evaluation of the programme was problematic, the success of the OLPC initiative in the US, Peru, Rwanda, and Tanzania was also likely constrained by poor implementation efforts, limited internet access in schools, poor support for repairs, non-child-directed delivery and lack of teacher training; all of which could have hindered children's learning with this technology (Hubber et al., 2016; Camfield et al., 2007). This highlights the importance of environmental factors in successfully deploying EdTech interventions in low-income contexts, which is enhanced when the children concerned are out of school and not receiving regular support from a trained teacher. It would also suggest that providing hardware and computer access with no suitable software and pedagogical content is unlikely to improve learning outcomes for out-of-school children (Sampson et al., 2019).

To establish what research already exists that combines high-technology and pedagogical content, a relevant literature search was conducted in December 2022 to find studies using mobile technology to improve foundational learning for out-of-school children in LMICs. Updated searches were conducted in June 2023 and February 2024. The inclusion criteria for this search were documents in the English language that report on the use of educational technology interventions to improve foundational literacy and/or numeracy skills in out-of-school children in LMICs. Inclusion was restricted to English documents for pragmatic reasons due to language restrictions. The search was limited to records between 2007 and 2024, chosen due to the rapid increase in the use of educational technology (e.g. Apple released the first iPhone in 2007; Apple, 2007). The World Bank classification of LMICs was used as guidance for whether a country should be included or not, last updated in 2022 (World Bank, 2022b). For this research, a child was defined as being between ages 6-17 years due to being school-age, and thus the target population when exploring research with 'out-of-school children'. This age group is the focus of Sustainable Development Goal 4 when referring to children and young people completing primary and secondary education (Webb et al., 2017; UNICEF, 2020). Finally, the working definition of mobile technology for inclusion was taken to be any portable electronic devices such as tablets, smartphones, and laptops (Fietzer & Chin, 2017). Low-technology options were not considered, and desktop computers were excluded due to their lack of portability, which is crucial in this context.

Fifteen academic databases were searched using a comprehensive search string, along with relevant research groups (e.g. EdTech Hub and The World Bank), non-governmental organisations (NGOs; e.g. UNESCO) and other grey literature (see Appendix A for the full search string). Studies investigating children receiving EdTech interventions as supplementary provision or children who were only out of school due to the COVID-19 pandemic school closures were excluded, as these are a different demographic to children unable to access school on a permanent basis. Five alternative technological initiatives were found to have already been developed and implemented with out-of-school children. Each of these initiatives took a slightly different approach to solving the fundamental problems, as described below, and summarised in Table 1.

Table 1

Implementation data extracted from the five studies identified in the literature search for EdTech interventions with out-of-school children

Study	**Participants** *Number,* *Age range,* *Country*	**Intervention**	**Intervention dosage**	**Infrastructure provided/required**	**Effect sizes** ***(as reported)***
Brezeal et al., 2016	40 4-11 years Ethiopia	Pilot study of the Curious Learning platform (literacy)	No direction regarding use.	Smartphone devices, Solar charging equipment, Training for adults in the villages.	N/A Only anecdotal evidence

Study	**Participants** *Number, Age range, Country*	**Intervention**	**Intervention dosage**	**Infrastructure provided/required**	**Effect sizes** ***(as reported)***
Orozco-Olvera & Rascon-Ramirez, 2023	9393 *(938 households which received smartphone)* 6-9 years Nigeria	Curious Learning platform (literacy) Apps used: Feed the Monster and Global Digital Library *Part of a wider intervention to reshape parental attitudes*	Five-day video intervention Smartphone provided at end of the 5 days for 12 months– no restrictions of use.	Video screenings targeting parental attitudes, Smartphone devices, Training for parents and target children, Solar chargers.	Not able to calculate Cohen's *d.* Reported: Literacy: z-score increased with an effect size of 0.462 ($p<0.01$). Numeracy: z-score increased with an effect size of 0.628 ($p<0.01$).

Study	**Participants** *Number, Age range, Country*	**Intervention**	**Intervention dosage**	**Infrastructure provided/required**	**Effect sizes** ***(as reported)***
Stubbé et al., 2016 *Pilot 1*	86 7-11 years Sudan	E-Learning Sudan (game-based numeracy app)	6-week intervention 5 times per week 45 minutes per day	Tablet technology (shared one per two children), Learning centres overseen by a facilitator running two sessions a day.	Not sufficient information to calculate effect sizes.
Stubbé et al., 2016, 2017 *Pilot 2*	916 7-9 years Sudan	E-Learning Sudan (numeracy)	6-month intervention 5 days per week 45 minutes per day	As above.	Not sufficient information to calculate Cohen's *d*. Reported: r = 0.85 for pre-post-test improvement in numeracy scores across the sample (p<.001).

Study	Participants *Number, Age range, Country*	Intervention	Intervention dosage	Infrastructure provided/required	Effect sizes *(as reported)*
XPRIZE, 2019	2041 7-11 years Tanzania	GLXP competition (see Chapter 2 for five apps used; literacy and numeracy)	15-month intervention – no restrictions of use.	Tablet technology (one per child), Solar charging stations, Village Mamas for facilitation/technical guidance.	King et al. (2019; impact report) reported effect sizes for the EGRA and EGMA subtests, ranging between 0.46-0.59 for literacy and 0.42-0.59 for numeracy.

As previously evidenced, programmes utilising handheld hardware, such as phones or tablets, have improved foundational learning outcomes in schools when used with appropriate learning software in low- and high-income settings. A novel approach developed in 2011 was the 'Curious Learning' initiative, using smartphones and a cloud-based backend to develop a platform of existing literacy-based open-source learning apps and distribute the product to marginalised children in overcrowded and under-resourced schools (both rural and urban) or out-of-school in remote villages (Gottwald et al., 2017; Brezeal et al., 2016). The scheme aims to reach 770 million people and currently spans 195 countries, providing content in 69 languages with over 75 learning apps from a variety of partners that enable children to develop their literacy and reading skills (Korin, 2021).

A pilot study was conducted in Ethiopia, deploying mobile phones with the Curious Learning platform to children in two remote villages that had no access to formal schooling but no real results. There were no appropriate digital activities or software targeted at the villagers' native dialect (Oromo), but their local government requested English deployment as it may improve employment opportunities for the children, so the devices were loaded with English literacy content, books, and videos, with educational content adjustable to the needs of the learner (Breazeal et al., 2016; Gottwald et al., 2017). All children between the ages of 4-11 years were provided with a device, with 40 children participating in total across the two villages, and solar charging equipment and training were provided to adults in the villages (Breazeal et al., 2016). The intervention objectives were for the illiterate children to develop technological familiarity, basic conceptual knowledge and vocabulary, early literacy skills and maintain sustained engagement.

It is not stated how long the intervention was intended to last, but due to funding issues, the researchers lost contact with the villages for one year, preventing the formal evaluation from taking place (Gottwald et al., 2017). Early Grade Reading Assessments

(EGRA) were conducted with six of the participants several years later to collect anecdotal evidence, with results demonstrating that the participants could accurately read text written in Oromo and answer comprehension questions, with two children accurately decoding 70% of the English words they were presented with (Gottwald et al., 2017). While these results are promising, the evaluation is limited and only based on six children's progress, which is insufficient to draw conclusions about the broad value of the Curious Learning programme in helping out-of-school children learn foundational skills.

More recently, the World Bank built on this research by conducting a cluster randomised trial testing two components of a five-day Curious Learning intervention with out-of-school children aged 6-9 years and their parents in 9393 homes in northern Nigeria (Orozco-Olvera & Rascon-Ramirez, 2023). Community video screenings were shown to half of the households to reshape parental attitudes to education, aiming to improve school attendance rates (with the other half forming a control group). Approximately a third of the children from the intervention group received a smartphone with the pre-loaded Curious Learning platform (Orozco-Olvera & Rascon-Ramirez, 2023). Two literacy apps were chosen for use on the basis that they can run offline, Feed The Monster (game-based learning) and Global Digital Library (reading resources), and were translated into the local language, Hausa. No learning instruction or restrictions were placed on the children; the goal was to provide increased and flexible access to literacy resources (Orozco-Olvera & Rascon-Ramirez, 2022).

After 12 months, children in the treatment group (video screenings) were 42% less likely to be out of school, indicating that the videos were successful in boosting attitudes and school enrolment, but there were no significant improvements to learning outcomes (World Bank, 2023). The combined intervention (video screenings and mobile provision) increased literacy skills by 0.46 standard deviations and maths skills by 0.63 standard deviations, measured by the Early Grade Reading and Maths Assessments (EGRA and EGMA; Orozco-Olvera & Rascon-Ramirez, 2022; World Bank, 2023). These large effects placed the intervention as the fifth highest scoring out of 74 learning interventions studied by the World Bank, and 0.5 standard deviation increases were equivalent to five years of instruction in the World Bank study schools (World Bank, 2023). There were also spillover effects for older siblings in the household, whose literacy and numeracy skills increased significantly, and a reduction of 13% in adolescent parenthood and a 14% reduction in early entry into the workforce, which are overwhelming problems in low-income countries (Orozco-Olvera & Rascon-Ramirez, 2022; World Health Organisation, 2023; Garcia & Fares, 2008).

While the findings are encouraging, several important questions still need to be addressed. Research has recommended that educational interventions must be implemented for a minimum of 12 weeks in order to assess its full potential; the current intervention period was only five days, which may be insufficient to draw reliable conclusions from (Higgins et al., 2012). EdTech interventions, in particular, should be implemented over a long test period to identify whether uptake of the intervention is due to the novelty effect of using tablets for the first time or sustained interest in the learning process (Huntington et al., 2023a).

Secondly, the Curious Learning platform provides access to a multitude of different learning apps rather than a specifically developed learning app for the targeted learner. Providing young children with a large choice of learning apps may overwhelm the children, leading to less concentrated learning time, and there is a wide range in the quality of available learning apps for children (Hirsh-Pasek et al., 2015; Kolak et al., 2021; Vaiopoulou et al., 2023). Furthermore, while the research showed that out-of-school children can learn using the platform and mobile devices, the evaluation of the programme broadly (as opposed to the individual apps) provides no information about the software features that do or do not facilitate learning, and thus would be helpful to use with these out-of-school learners.

Despite the remaining questions, these broadly positive findings across the two studies conducted in Ethiopia and Nigeria show promising evidence to suggest that a home-based smartphone implementation pre-loaded with the Curious Learning platform can substantially improve foundational literacy and numeracy skills for out-of-school children, and that these skills are sustained for the following year, making mobile technology with appropriate software a potential low-cost mitigator for the learning crisis experienced in LMIC.

Limited research has been conducted on the effectiveness of learning app software with out-of-school children that specifically uses tablet-based technology, which is the technology evaluated throughout this thesis. This could be due to many reasons, such as the anticipated cost of distributing hardware (as existing penetration of tablets is so low within LMIC out-of-school communities), the difficulty of accessing the population for research, and the lack of research demonstrating its effectiveness in the specific context (Kukulska-Hulme, 2023; World Bank, 2018; Tauson & Stannard, 2018). The few studies that have examined foundational learning with tablet technology in LMIC contexts with out-of-school children are described below.

The first tablet-based mathematics programme tested with a remote, out-of-school population without access to learning materials was E-Learning Sudan (2012-2015). Game-based tablet technology with video instructions was used to autonomously teach early mathematics skills, such as counting, addition, subtraction, and place-value concepts to primary-aged children in remote villages (Stubbé et al., 2016). Two quasi-experimental pilot studies were conducted, with sessions taking place in built-for-purpose learning centres without teaching support but overseen by a facilitator. All tablets were shared between two children due to two separate sessions taking place per day, and tablets were not accessible for use between sessions. Learning improvements were measured using local school-based tests to assess Grade level performance.

In the first pilot study, 86 out-of-school children aged 7-11 from four remote communities in Sudan took part in the experiment (19 in the control group receiving no education) and played the tablet-based game five times a week, 45 minutes per day, for a period of six weeks. Results found that the experimental group demonstrated significantly greater improvement in maths test scores (pre-post test) compared to the control group. There was also a large effect size between the experimental and control group post-test scores (Cohen's d = 1.16), indicating that the experimental condition influenced a greater acquisition of foundational maths skills.

In the second pilot study, participants were 916 children aged 7-9 years from 29 communities across three states in Sudan (White Nile, North Kordofan and Gedaref; Stubbé et al., 2016). The experimental group again played the e-learning game for 45 minutes per day, five days per week, but for a longer experimental period of six months. The control group (325 children from 10 communities across the three states) were enrolled in informal education, receiving two 45-minute mathematics lessons a day taught by a teacher in out-of-school centres. All three experimental groups showed a significant improvement in scores from pre- to post-test, with a high effect size of r=0.85, demonstrating that the tablet-based learning intervention is successful at improving basic mathematics skills with out-of-school children. However, there was no significant difference between the experimental and control conditions for children in the North Kordofan state (pre-test, post-test, or score increase). The authors concluded this was a positive finding, as the control group received twice as much learning instruction (twice a day vs once a day) yet still failed to outperform the experimental group (Stubbé et al., 2017). Results showed that for the White Nile state, the control group outperformed the experimental group with a significant difference in score increase from preto post-test (low-medium effect size of r=0.24), suggesting that the additional instruction the control group received was modestly effective in enhancing foundational mathematics skills. These findings highlight the importance of instruction time for improving learning outcomes but also suggest that contextual factors (such as infrastructure and community) may influence learning for out-of-school children, as demonstrated by the differences in outcomes between regions when looking at the relationship between the control and experimental groups.

These two pilots provided an insightful exploration of the effectiveness of tablet-based technology for out-of-school learning, but there were multiple methodological problems with the second pilot that threatened the reliability of the study findings. Data collection issues meant that the pre-and post-tests were not conducted at the planned times; the Gedaref pre-test was taken two months after the beginning of the intervention, so all 100 control group children were excluded from analyses (Stubbé et al., 2016). The post-tests were taken 3 (North Kordafen) and 6 months (White Nile) after the intervention instead of the planned six weeks for the other two states. There were also issues with logged data; two communities had no data collected and matching logged data to test data was difficult, leading to loss of data (23 files) and matching data based on name instead of case numbers. These identified challenges suggest that the results from the intervention should be interpreted with caution, and further research is needed to establish the effectiveness of the E-Learning platform. The problems faced during these pilot programmes further highlight the difficulty of implementing and evaluating EdTech interventions with out-of-school children, emphasising the need to carefully plan the delivery and data collection processes with this hard-to-reach population, as they present additional, unique challenges. Furthermore, as with the radio-based learning programmes, the scheduled sessions could provide a substantial barrier to educational access for girls and children with higher household commitments. A key advantage of EdTech for out-of-school children is its flexibility of use, which is eliminated during this structured programme, particularly as the tablets are not available between the learning sessions for the children to use freely (Tauson & Stannard, 2018; Stubbé et al., 2017).

Another initiative that tested the ability of tablet-based technology to improve foundational learning skills for out-of-school children is the Global Learning XPRIZE (GLXP) competition, a unique field trial that took place between 2017 and 2019 (XPRIZE, 2019). Global, multi-disciplinary teams were challenged to develop open-source, scalable learning software to improve foundational reading, writing and numeracy skills. Five finalist teams field-tested their app (deployed on tablet technology) with 2041 illiterate out-of-school children aged 7-11 years from rural villages in Tanzania over a 15-month period. All five teams demonstrated significant core improvements in literacy, maths and writing skills over the duration of the trial, with children who received instruction from the two winning apps, onebillion and KitKit School, achieving the greatest overall proficiency gains (XPRIZE, 2019). As the competition was successful in improving learning outcomes and the field trial compared five distinct learning apps with the same population, it provided a unique case study opportunity, which this thesis research is based upon, by providing an in-depth secondary analysis of the GLXP findings (see Chapter 2 for further details of the trial, software, and results).

In summary, research conducted with long-term out-of-school children is scarce, and the research quality is generally poor; hence, there is currently a distinct lack of rigorous, robust evaluation due to the methodological challenges faced. It has been demonstrated here that programmes only offering hardware solutions produced mixed results for educational outcomes, while initiatives that provided hardware with curriculum-driven software installed offered greater promise in fostering foundational learning skills with this marginalised demographic.

Many of the out-of-school initiatives involved setting up learning centres where children gathered daily to receive the intervention that was supported by adult facilitators (as in E-Learning Sudan; Stubbé et al., 2016 and the Taonga Market programme; Education Development Center, 2020). Learning centres may be problematic in more rural contexts, due to scarce resources and small communities making it difficult to obtain volunteers to facilitate the learning sessions (Mokoena, 2015; Wright & Plasterer, 2012). Furthermore, as the centres have to be staffed, there would likely have to be scheduled session times, which presents the same barrier to accessibility for girls that the fixed radio broadcasts do. Therefore, further research is needed to investigate whether children can learn autonomously using tablet-based technology in their home environments, as this could minimise the resources needed for the learning process.

Little is yet known about the circumstances in which out-of-school children learn autonomously in the home environment without an organised structure. Research is needed to identify the child-level and contextual factors that are most important in determining the effectiveness of EdTech interventions, particularly those with tablet-based technology implemented in rural, low-income communities, which will be explored in Chapter 5. This knowledge would then significantly contribute to the optimisation of future programmes with this population.

Consideration of gender inequalities

As girls were identified as a high-risk group impacted by the education crisis in low-income countries, it is important to determine what gender differences currently exist, the importance of supporting girls' education and the potential role EdTech may have in reducing existing inequalities.

Out-of-school differences. In 2021, approximately 245 million children and youth were out of school, comprising 119 million girls and 126 million boys (UNESCO, 2022b).

This data initially suggests a relatively balanced gender distribution among out-of-school children. However, there are significant regional discrepancies, particularly in LMIC and notably in sub-Saharan Africa, where the female out-of-school rate is 4.2% higher than for boys at the primary level and as high as 20% at the secondary level (UNESCO, 2022a, UNESCO, 2022c). Despite the significant decrease in gender gaps in education enrolment globally, sub-Saharan Africa worryingly lags behind, exhibiting the widest gap to the detriment of girls' education and highlighting the need to address these imbalances (UNESCO, 2022c). Less than two-thirds of girls enrolled in primary school in low-income countries completed their schooling (UNESCO, 2020). School retention falls substantially for girls due to a multitude of socioeconomic factors such as poverty, marriage, pregnancy and having traditional gender roles to fulfil within their households (UK Aid, 2021; Awinia, 2019). As a result, there are significant gender disparities in learning outcomes at primary school, which widen further after the transition to secondary school, with girls generally receiving lower examination results than boys (Awinia, 2019; Al-Samarrai & Tamagnan, 2019). This exacerbates the dropout rates further, with a 1:2 girl-to-boy ratio by the time students reach upper secondary (Pezzulo et al., 2022). This gender inequality is likely to get substantially worse for girls in low-income countries, as they were more likely to have missed out on foundational learning during the pandemic than their male counterparts (Conto et al., 2020).

Interestingly, while girls generally outperform boys in reading, boys tend to excel in mathematics during foundational education (UNESCO, 2022a). Girls' performance in mathematics improves in more gender-equal societies, as seen in primary education in LMICs, where opportunities are more equally distributed among genders (Guiso et al., 2008). Therefore, it is imperative that learning opportunities are developed with both genders in consideration to eliminate barriers that would prevent either gender from learning effectively across domains.

The discrepancies between male and female school enrolment and educational outcomes are widely recognised, leading to a consensus among governments and charities on the urgency to address the global challenges that inequalities create. In response to this crisis, the UK became the largest global funder of girls' education in 2012 by launching the Girls' Education Challenge (GEC) with the Department for International Development (DFID) pledging a 12-year commitment to researching and providing quality education to marginalised girls globally (GOV.UK, 2013). This initiative, representing a commitment of £855 million over 12 years, underscores the serious investment in and acknowledgement of the importance of educating girls worldwide (DFID, 2016; Reilly, 2023). The GEC's approach has used a variety of projects to directly provide high-quality education to over a million disadvantaged girls. Some of these key projects included EdTech interventions, highlighting the pivotal role of such initiatives in enhancing girls' educational experiences (DFID, 2018).

The importance of prioritising girls' education is due to its multi-level benefits for the individual and their family, community, and society. Research has shown that increasing schooling is likely to positively affect girls at an individual level (Musomi & Swadener, 2017), including increasing the potential for higher earnings, where one year of extra schooling can equate to a 10-20% increase in the salary they are capable of getting (Psacharopoulos & Patrinos, 2002; Tikly & Barrett, 2011). However, conflicting research has suggested that whilst increasing girls' educational attainment and skills is critical to attracting income-generating opportunities, these educational gains have not translated into better financial outcomes for most women (UN Women, 2015, p.69). There could be many reasons for this, such as the inflexibility of jobs around childcare provision and the disproportionate responsibility women have for unpaid care and domestic tasks (UN Women, 2015). Despite this, governments must provide the education necessary to empower girls and allow them access to opportunities that arise.

Another benefit is that girls who receive formal schooling throughout the transition to secondary school are more likely to delay early parenthood and marriage, with differences of 10% found in fertility age when only completing one additional year (Ferre, 2009). Marriage in some marginalised countries is common in children as young as 14, and one in five girls get married under the age of 18 worldwide (UNICEF, 2018). Childhood marriage threatens the quality of life that girls lead, with a heightened likelihood of domestic violence, isolation from families and their community and school dropout (Lloyd et al., 2007; Erulkar & Muthengi, 2009). These implications would have a substantial cost to the societal economy and provide evidence that the lack of education can provide a vicious cycle where, if you receive inadequate schooling, it may lead to childhood marriage, heightening the likelihood of school dropout. The potential impacts of early marriage and childbirth on both the individual and society are clear, and they highlight the significance of addressing gender-based learning inequalities.

The benefits extend to families and communities, where educated women contribute to poverty reduction and community development. Their involvement in the workforce and community roles is critical, especially in areas with a labour scarcity. This contribution enhances community structures and fosters economic growth and sustainable lifestyles (Heath & Jayachandran, 2017; Müller, 2019; Ghuman & Lloyd, 2010; Roseland, 2012). Moreover, the intergenerational impact of educating girls is evident, with educated mothers more likely to raise healthier, better-educated children, creating a positive cycle of empowerment and development (Grépin & Bharadwaj, 2015; Gakidou et al., 2010; Aslam & Kingdon, 2012; Behrman et al., 2009; Heath & Jayachandran, 2017). Education quality and actual attainment were cited as the most critical factors, rather than the length of schooling experience (Andrabi et al., 2012).

At the societal level, educating girls is a catalyst for economic growth, potentially fostering up to 20% growth with each additional year of schooling (Hanushek & Woessman, 2011). This prospect is compelling for governments and policymakers to invest in female education, recognising that the returns extend far beyond mere financial gains (DFID, 2020). Ultimately, the benefits of educating girls are extensive, reducing violence, child marriage, and pregnancy rates while boosting employment opportunities and economic growth and addressing critical social issues like domestic violence and harmful cultural practices like female genital mutilation (FGM; UNICEF, 2018; Lloyd et al., 2007; Erulkar & Muthengi, 2009).

Understanding gender differences in learning is crucial to help educators and policymakers address the current inequalities and provide tailored solutions for both genders, and thus the exploration of gender is a key theme throughout this thesis. EdTech may be a solution to mitigate gender-based learning differences, as research has suggested that it may enable access to education due to the ability to deliver foundational learning at scale and despite distances (Samuels et al., 2022). Educational technologies may be able to promote gender-equitable outcomes, even without an explicit focus on gender equality (Evans & Yuan, 2022). Growing evidence suggests that if equity and inclusivity are considered when designing EdTech software, these tools can promote higher engagement for girls than boys, reinforcing the potential that EdTech may have to address educational gender disparities (Webb et al., 2020).

Research by Pitchford et al. (2019) supports the potential use of EdTech to bridge the educational gap between genders. They showed that in Malawi, the implementation of a 14-month app-based intervention (designed by onebillion) delivering eighteen 30-minute mathematics sessions over the first year of schooling in Malawi prevented the gender differences emerging in numeracy that typically advantage boys in traditional schooling. After the EdTech intervention, the girls in the intervention group improved in numeracy attainment significantly more than the control group, evidenced by a relatively large effect size (Cohen's $d = 0.619$). Furthermore, in the intervention group, no significant differences were found between numeracy improvement for girls and boys ($\eta 2 = 0.001$). However, in the control group, boys achieved significantly higher attainment than girls ($\eta 2 = 0.012$). This difference in findings between the intervention and control groups indicates that the EdTech solution facilitated the girls to catch up to the boys in numeracy, demonstrating that EdTech may have a clear advantage for girls' numeracy attainment over traditional schooling practices. Furthermore, no significant differences were found in literacy and numeracy learning improvement in the out-of-school initiatives described previously that explored gender differences (Stubbé et al., 2016; Orozco-Olvera & Rascon-Ramirez, 2023). These findings suggest that EdTech may be highly valuable for girls if educational access is provided.

While these results are promising, not much is currently known about the true potential that EdTech has to close the gender discrepancies found in traditional methods of foundational learning across domains. It is clear that EdTech may mitigate gender differences, but research needs to explore what elements of EdTech are responsible for bridging the educational gap, and if these operate across both literacy and numeracy domains, allowing educators to adapt educational software accordingly to support girls' foundational learning. Popular app features and their impact on both genders and domains (literacy and numeracy) will be investigated in Chapter 4, as gender discrepancies remain a crucial issue in LMIC and are relatively underexplored. However, it is essential to recognise that even if EdTech is successful for both genders, girls may not have the same access to EdTech opportunities and benefits as boys, particularly in LMIC (Crompton et al., 2021c). As previously mentioned, many barriers to access are specific to girls, such as early marriage, pregnancy, and household responsibilities, so it is vital to establish if these gender-specific barriers are still impactful even when a flexible, autonomous learning solution is implemented.

Children with Special Educational Needs and Difficulties (SEND)

Another group identified as being among those most affected by the global learning crisis is children with special educational needs and difficulties. This thesis did not specifically focus on children with SEND, as the case study it is based on - the GLXP competition - did not establish whether the participants had any additional needs. However, there are nearly 240 million children with disabilities globally, and this demographic is 49% more likely to have never attended school (UNICEF, 2021b). There are at least 29 million children with documented disabilities across Eastern and Southern Africa, with the true number likely much higher due to difficulties with diagnosis, and fewer than 10% of disabled children in Africa attend school (Moumen, 2023). Accordingly, there is likely to be a high proportion of children with SEND amongst the out-of-school population. This may impact the relative effectiveness of EdTech interventions delivered directly to the out-of-school community compared to school-based interventions.

Research has suggested that tablet-based technology may facilitate foundational learning for children with SEND due to their portability, ease of use, accessibility features, and elimination of tools that need higher dexterity skills (e.g. a computer mouse; Kucirkova, 2014; Ryan, 2016; Coflan & Kaye, 2020). However, existing evidence is predominantly exploratory or conducted in developed, high-income countries (Pitchford et al., 2018; Shah, 2011; Badilla-Quintana et al., 2022). A recent systematic review evaluated the existing evidence of the use of EdTech to support learners with disabilities in low-income countries and found that only seven studies (out of 51 total) looked at the impact of the EdTech implementation on academic learning outcomes, such as literacy and numeracy skills (Lynch et al., 2024). Those studies did not attempt to measure or explain engagement or the changes in learning outcomes, leaving researchers to wonder if the novelty effect of introducing new technology was driving engagement.

Five studies focused broadly on SEND rather than on a specific population such as children with visual impairments or deaf learners (Lynch et al., 2024). Of those five studies, one focused on accessible technology rather than specialist assistive technologies, which may provide accessibility to mainstream content for children with SEND but is not specifically designed for this population (Coflan & Kaye, 2020). The study was observational, implementing tablet technology with onebillion numeracy software for 33 children attending a SEND unit in two state primary schools in Malawi (Pitchford et al., 2018). The intervention was delivered daily as standard maths practice for children with SEND. The rate of progress through the topics was calculated to assess learning gains and then correlated with the teacher ratings of disability severity and engagement with the learning app. Results showed that SEND pupils could interact with the apps and learn basic numeracy skills but at half the speed of mainstream pupils (approximately four hours per topic vs two hours), and the extent of the child's disability significantly predicted their progress rate (Pitchford et al., 2018). This would suggest that tablet technology with mainstream, curriculum-aligned learning software could improve foundational learning outcomes for children with SEND, although larger, rigorous quantitative studies need to be conducted with similar populations for more robust conclusions to be made. Research also needs to investigate which technologies, learning features and contexts facilitate learning gains rather than just measuring improvement in outcomes (Lynch et al., 2024).

There is seemingly no documented research testing the impact of an app-based EdTech intervention with out-of-school children with SEND, which is not surprising considering the general scarcity of EdTech learning interventions with out-of-school populations. As already established, children with SEND in low-income countries are likely to be out of school, but without a formal diagnosis, there is no way of knowing which children have SEND. As children with SEND often struggle to learn at the same rate as their mainstream peers using EdTech in school, it is likely that the same might be seen with out-of-school children (Pitchford et al., 2018). Therefore, it would be valuable to investigate whether relevant child and community-level contextual factors can predict whether an out-of-school child will respond (or not) to an EdTech solution. This could help educators evaluate whether a community is likely to benefit from an EdTech intervention. If a community has a large proportion of predicted non-responders, it may suggest a high SEND population, and developers could target these children using app design or implementation strategies identified as being successful in facilitating learning for children with SEND.

2) The mechanisms through which out-of-school children may acquire foundational skills using tablet-based technology and app-based learning software.

While mobile learning apps are becoming increasingly prominent in education, there are very few models aimed at understanding how digital learning manifests and what app features encourage digital learning to translate into successful learning outcomes, especially in remote out-of-school settings (Qureshi et al., 2020; Kim et al., 2021). Determining which mobile app features facilitate learning improvements is crucial as the expansive market includes many educational apps that are poorly designed, with low educational value, and lack features that are known to support children's learning (Larkin, 2015; Kanders et al., 2022; Callaghan & Reich, 2018). When an app has been shown to be effective, the focus tends to be on learning outcomes rather than app design, preventing developers from obtaining sufficient descriptions of successful mechanics and thus replicating the aspects that have driven positive outcomes (Alam & Dubé, 2022).

One model of digital learning that has been used to evaluate educational apps is the Four Pillars of Learning framework by Hirsh-Pasek et al. (2015). This focuses on four app design aspects that evidence has shown may positively affect a child's learning process:

1. Active learning – Does the app require thought, attention and intellectual effort?
2. Engagement in the learning process – Is the app interactive? Does it engage rather than distract?
3. Meaningful learning – Is the app meaningful to everyday experiences? Is the app contextually appropriate for the children?
4. Social interaction – To what degree can children interact with (i) Caregivers and (ii) Characters within the app?

This framework has been applied to school-based app interventions (e.g. Outhwaite et al., 2019b) and used to evaluate the quality of learning apps freely available on digital app stores (Meyer et al., 2021). The principles have yet to be applied to an out-of-school learning context, so it is vital to consider how suitable each of the four pillars of learning may be for this demographic, and what the unique challenges the out-of-school context may provide for this model.

Designing apps that engage children in learning is widely considered successful in enhancing learning outcomes. This process should involve creating an interactive environment for the child, providing a multi-sensory learning experience involving animation, and personalised user choices (Garcia, 2019; Department for Education, 2019; Giraldo et al., 2021). A careful balance needs to be created with young learners as a lower attentional capacity may mean that anything other than basic sounds and animations is more distracting than usefully interactive due to the heightened cognitive load (Alam & Dubé, 2022; Willoughby et al., 2015). Furthermore, keeping children engaged with a learning app in a low-income, unstructured environment may be difficult. A review testing out-of-school learning provision in low-income countries during the COVID pandemic indicated that having sufficient resources, such as a stable electricity source, was crucial for sustained engagement in a learning app (Nicolai et al., 2023). This is likely so that the learning process is consistent and not disturbed by technological issues. Furthermore, if the children do not know how to use the technology, they may struggle to engage with the app, so out-of-school programmes would need to involve training for the children to feel comfortable with the technology (Islam et al., 2022). These examples suggest that with out-of-school children, engagement may be less about the quality of the in-app animations and more about the set-up process and quality of the infrastructure that the EdTech programme provides.

Research also supports the use of social interaction within an app (Berkowitz et al., 2015). When parents are trying to choose high-quality apps, the UK's Department for Education suggests that a learning app should provide both in-app interactions with virtual characters and external interactions with others, such as a parent and teacher, for guided support (Department for Education, 2019). This should then facilitate valuable and rich discussions around learning material (Berkowitz et al., 2015; Department for Education, 2019). Research with out-of-school children has also attributed children's engagement to receiving scaffolded support from motivated teachers (when during COVID) and parents or caregivers who nudge children to engage and persevere with the app-based learning material (Giraldo et al., 2021; Islam et al., 2022). This suggests that learning apps that prioritise social interaction within an app may also be successful for out-of-school children, although it does contradict the concept that children may be able to learn truly autonomously with tablet-based technology.

Studies suggest that any interventions and corresponding activities need to be grounded in the curriculum to be successfully implemented within a school setting (Hennessy & London, 2013; Pitchford, 2015; Berkowitz et al., 2015), as this is the existing pedagogy proven by consistent evidence to foster learning in children (Osborne & Hennessy, 2003). The learning content should also teach topics appropriate for their target age to stay aligned with their understanding, which should be possible if successfully aligned with the curriculum (Berkowitz et al., 2015). While a sensible suggestion for best practice, keeping content aligned with curricula may prove challenging when implementing a specific intervention over several different countries due to the substantial differences in curriculum between areas and contexts. Furthermore, in low-income countries where children are poorly educated and potentially illiterate, targeting content based on age may be problematic, as the learning stage of the children would be substantially behind that expected of their age (Lewin & Sabates, 2012). Aligning content to the curriculum could be even more challenging for out-of-school learning provision. There are likely to be huge variations in children's level of prior knowledge and learning pace, particularly when the number of children with undiagnosed SEND is assumed to be high (p.50-51).

Personalisation of an app has the most substantial evidence base to support its effectiveness in improving children's foundational learning outcomes and is widely recommended for the implementation and choice of high-quality educational learning apps for developers and parents alike (Porter, 2018; Department for Education, 2019; Outhwaite et al., 2023). While there are varying definitions of what constitutes 'personalised learning' in an EdTech context, there is a consensus that the software should be learner-centred, flexible, and tailored to individual strengths to promote mastery of skills (Gro, 2017; Vanbecelaere et al., 2020a). The adaptive nature of personalised technology allows users to have a unique learning experience adjusted to their age, knowledge, context, preferences and learning styles (Giraldo et al., 2021; Major et al., 2021). This flexibility could be particularly effective for out-of-school children, as it helps to alleviate potential concerns about children starting from varied stages of the curriculum and progressing at different paces. Adaptive software also provides the opportunity to empower children by allowing them to choose how, when and what they learn, within reason, and it responds to established learning routines accordingly (Major et al., 2021).

An element of personalisation considered crucial for a high-quality learning app is the provision of instant feedback based on the input of the learner (Giraldo et al., 2021; Garcia, 2019; Department for Education, 2019). Feedback provides learners with differentiated performance and progress information, shaping their responses, behaviour and learning process (Alam & Dubé, 2022). Recent research has shown that explanatory and motivational feedback is required for learning improvements in literacy and numeracy, as part of a personalised learning approach with levelling (Vanbecelaere et al., 2020b; Outhwaite et al., 2023). However, feedback becomes less impactful after the tasks are completed for the first time (Callaghan & Reich, 2021). In an out-of-school setting, feedback is potentially more important. This is because the children may receive motivational feedback from a caregiver, but are unlikely to receive explanatory feedback, due to caregivers having low levels of education themselves (Islam et al., 2022).

Personalisation, in combination with feedback, can be beneficial for children in low-income countries, as an educational provision that adapts to the level of the learner can be cost-effective and help mitigate the negative impact of large class sizes, and limited numbers of high-quality teaching staff (Jordan et al., 2021; Kishore & Shah, 2019). Personalised, adaptive software also offers the potential to support autonomous or ‘self-led’ learning, increasing accessibility to children who may otherwise have no access to schooling or learning resources (Major et al., 2021). A recent meta-analysis showed technology-supported personalised learning to have a significant positive effect on learning in LMIC, with an effect size of 0.18 (Major et al., 2021). The Education Endowment Foundation considers this effect size equivalent to two months of additional progress, albeit in a UK school setting (Major et al., 2021; Education Endowment Foundation (EEF); 2023).

Whilst the features above are among the most prominent, additional research has compared other app features to traditional learning tools, such as gamification/play-based learning, direct instruction, task structure, meaningful learning and accessibility (Ku et al., 2014; Alam & Dubé, 2022; Department of Education, 2019; Outhwaite et al., 2023; Haßler et al., 2015). However, existing research primarily focuses on Westernised settings, so it is unclear whether the features would also be beneficial for learning in a low-income setting, particularly with out-of-school children (e.g. Outhwaite et al., 2023; Kolak et al., 2021; Papadakis et al., 2017).

Most EdTech intervention research focuses on learning outcomes, but few attempt to explain or measure the changes in learning outcomes after implementing the intervention (Jordan et al., 2021). As a result, outcomes could result from a novelty effect, where learners react positively because the learning method is new and exciting (Hew & Cheung, 2013). This is especially pertinent for LMICs because the penetration of smartphones and other handheld devices is extremely low, indicating that children will likely have little to no prior experience with handheld devices such as tablets (for example, Tanzania's smartphone penetration is 32% and tablets less than 1%; Lamtey, 2024). Further robust research has been called to evaluate app design and its impact in environments, conditions, and contexts other than Westernised countries and public schools, with different groups of learners across different domains (Lynch et al., 2024; Outhwaite et al., 2023). In response to this call for research, this thesis examined the impact of app features after intervention with out-of-school children in a rural, low-income country environment (Chapter 4), including an exploration of gender differences across learning domains (literacy, numeracy) and the effectiveness of these app features on foundational learning (Chapter 5).

Challenges with implementation and infrastructure

To effectively integrate an EdTech intervention into children's education, it is imperative to first understand the infrastructural barriers and challenges involved and then determine the conditions for successful implementation within school and out-of-school settings. Gulliford et al. (2021) suggested that while app features are important, successful implementation also relies on multiple child-level factors, the facilitator, and the learning environment. If not carefully designed, delivered, and used without considering the specific context, introducing an EdTech learning programme risks exacerbating existing learning inequalities rather than improving them (Allison, 2023).

While many effective practices identified are taken from EdTech research conducted in the high-income Global North, best practices are generally considered universal, based on the principle that high-quality teaching practices are beneficial across different contexts (Burns, 2022). When evaluating the implementation of the maths-based onebillion learning app in eleven UK schools, Outhwaite et al. (2019) identified four key themes from observations and interviews with participating teachers that were believed to contribute to effective intervention implementation - teacher support, teacher supervision, implementation quality and established routine. Established routine significantly impacted children's learning outcomes with the apps, predicting 41% of the observed variance. This indicates that having a clear, well-established routine in the classroom is crucial to EdTech integration in a school environment. This suggests that, for out-of-school children, when there is no funding for a centralised learning centre, having an informal set-up within communities may be helpful to motivate children's learning when they are out of school, even if it is just a specific outdoor spot for groups of children to meet and work together each day. However, for out-of-school settings, it has been previously mentioned that having a scheduled time and dedicated space to work on the tablets may be problematic for certain groups of learners, particularly girls, who have to balance other obligations, such as household responsibilities, and therefore struggle to access educational opportunities (Nicolai et al., 2023). Thus, the impact of an established routine for app-based learning with out-of-school children requires further investigation and careful consideration in relation to different groups of learners.

Another fundamental consideration when implementing EdTech in school environments is the need for technological support and adequate training provision for all teachers involved (de Oliveira, 2014; Haßler et al., 2015). Without this training, teachers are less likely to implement the interventions successfully, are going to hold less positive opinions about these changes and potentially feel stressed by them, which can lead to slower uptake of technological advances and poor implementation processes (Tom, 2018; Miglani & Burch, 2019). Therefore, research suggests that sufficient time, money, and resources should be allocated to implement any such intervention or change in teaching style to ensure teachers and staff are trained and supported appropriately. However, a recent study implemented the Headsprout Reading Programme to 269 learners in 22 Welsh primary schools over a 23-week implementation period. They found that providing an ongoing support model (including school visits, feedback, modelling, and practical advice) to staff delivering the programme did not affect the quality of delivery or learning outcomes for reading (Roberts-Tyler et al., 2023). This contradicts previous research, as it indicates that only minimal implementation training and ongoing technical support are necessary for high-quality EdTech provision (Roberts-Tyler et al., 2023; Naylor & Gorgen, 2020). However, ongoing implementation support was shown to make a noticeable difference for the teachers when they were assisting struggling learners who require additional support (Roberts-Tyler et al., 2023). This indicates that the specific learners being targeted should be considered when deciding what level of ongoing support should be provided. In an out-of-school context, where the majority of children would be classed as struggling learners, adult facilitators (whether formal or informal) should be appropriately briefed and receive initial training on how to support learners in engaging with the technology (UNESCO, 2021; UNICEF, 2022b). Regular contact opportunities could then be provided to monitor whether they need further support throughout the implementation period.

Whilst many implementation challenges are universal, there are additional infrastructural challenges in developing countries that must be considered. These include inefficient use of funds, obsolete technology and connectivity, chronically low staffing levels and a need for more organisation regarding procedure, policy, and communication (Adam et al., 2016; Nicolai et al., 2023). Within low-income contexts, even standard education instruction is highly dependent on weather, child labour requirements (e.g. harvest season), and teacher availability. These variables can highly impact the level of education a child receives during a specified period (Pitchford, 2015). These factors may restrict how teacher-led EdTech can be implemented successfully, particularly when still in the research context and the stringent conditions required for a randomised control trial (RCT). Higher levels of flexibility and contingency measures may be required for the successful implementation of an EdTech intervention within a developing country. The infrastructural obstacles are also likely contributing to the significant methodological challenges faced in some existing out-of-school initiatives (e.g. Stubbé et al., 2016). To identify which infrastructural challenges are most pertinent to the EdTech learning process for out-of-school children, it was important to evaluate whether contextual factors such as the children's prior experience with technology, support from community members and access to electricity in the household are contributing predictors for improving learning outcomes.

Another critical issue for implementing all educational programmes, particularly EdTech interventions, is cost, with governments and educators having to carefully prioritise where to invest their time and resources (Beeharry, 2021; Nicolai et al., 2023). The cost of hardware has proven to be a significant barrier in many low-income countries, with governments generally preferring to spend their budgets on tangible assets, such as building schools and hiring teachers (Passey et al., 2016; Kaguo, 2011). However, once hardware has been invested, adaptive learning software is considered a good purchase for low-income countries (World Bank et al., 2020). Similarly, research has shown that EdTech can be more cost-effective than textbooks over time, but perceptions about what appropriate learning materials are provide a barrier towards embracing and implementing innovative, technological learning methods (Tembey et al., 2021). Providing access to technology-based learning methods could be a highly cost-effective solution to reach children in the most remote areas without traditional resources (Allier-Gagneur & Coflan, 2020). However, to effectively implement EdTech solutions within this population, it may be crucial to first explore perceptions towards the technology by interviewing individuals that have experience in implementing EdTech with out-of-school children. Obtaining informed opinions about how receptive governments, children and surrounding communities are to an EdTech intervention will help educators to approach the sensitization period carefully. The goal would be to instil a positive attitude within governments and communities so that they feel that the programme is worthy of their attention, effort, and financial commitment.

Edtech deployment during the COVID-19 pandemic

The COVID-19 pandemic led to global school closures that affected 1.6 billion students from 180 countries (Global Humanitarian Overview, 2022), leading to an educational crisis where learning provision had to be adapted to continue in the home environment to avoid maximum disruption to the learning process for young children and adolescents. There was a significant shift towards remote education, with children accustomed to traditional schooling learning from home in less-than-optimal circumstances (Cortés-Albornoz et al., 2023). While unprecedented, this comprehensive movement to online learning platforms has provided unique insights into how EdTech can enhance and replace traditional learning methods when needed for children who are usually in school (Ma et al., 2020). It also inspired researchers to use the opportunity to test EdTech interventions as substitutions to regular instruction and establish their success. The lessons learnt from the deployment of EdTech during the global COVID-19 pandemic can inform future policy and practice in this rapidly developing field (Bettinger et al., 2020; Ma et al., 2020).

Although EdTech substitutions for regular instruction have been shown to positively impact foundational learning outcomes during the pandemic, the long-term effectiveness of learning programmes is still being evaluated in many settings (Muñoz-Najar et al., 2021; Nicolai et al., 2023). At a country level, evidence on the effectiveness of remote learning is mixed (Muñoz-Najar et al., 2021). Research from the US, Brazil, and the Netherlands suggested that students only learnt 60-80% of the minimum expected knowledge for maths and reading following a period of school closures, even if the children had an advanced technological set-up in the home environment (Muñoz-Najar et al., 2021; Engzell et al., 2020). On the other hand, evaluations in Australia and Uruguay have shown that test scores are consistent between past academic years and long periods of remote learning during the pandemic (Muñoz-Najar et al., 2021).

Furthermore, understanding the true effect of remote learning interventions during the COVID-19 pandemic and discerning their effectiveness in low-income countries is challenging due to the substantial bias in the existing literature (Crompton et al., 2021a). A review into emergency remote education highlighted that only three percent of studies focused on low-income countries, despite these populations being most in need of extra support amidst a learning crisis (Crompton et al., 2021b). However, an Interactive Voice Response mobile intervention in Bangladesh, delivering 60 audio lessons over 15 weeks to 1700 primary school children, showed significant improvements across literacy and numeracy learning outcomes, with literacy scores improving by 0.66 standard deviations (English) and 0.62 standard deviations (Bangla) and 0.56 standard deviations in numeracy, relative to children in the control group (Hassan et al., 2021). It was also found that the children from low-income families with the lowest baseline scores had the highest improvements by endline (with a 1% significance level). This finding is substantiated by other low-technology distance learning research (Hassan et al., 2021; Islam, 2022).

A recent review evaluated ten low-technology interventions conducted in LMICs during the pandemic and found that there were mixed results on the effectiveness of the intervention on foundational learning outcomes (Nicolai et al., 2023). While these were low-technology solutions, rather than tablet-based, this provides further evidence to suggest that children from low-income countries could learn from a remote intervention and indicate that the most marginalised children in society may benefit most from EdTech solutions.

As this thesis research is focused on the use of tablet technology specifically, the success of any tablet-based interventions in low-income countries throughout the COVID-19 pandemic must also be considered. One tablet-based initiative was the activity-based Sekolah Enuma Indonesia programme, developed by the EdTech company Enuma. Enuma also designed the game-based app KitKit School, which was one of the two winning apps in the GLXP competition (XPRIZE, 2019; Enuma, 2021). Pre-loaded tablets designed to improve literacy, numeracy, and additional language (English) skills were implemented over a three-month test period during the pandemic in 2021 with 586 children aged 4-9 years. On average, children played with the tablets at community sites for an hour a day, five times per week, with the sessions supervised by a local facilitator and multiple sessions per day meaning that the children could not use the technology at home (Enuma, 2021). One community (Medan) tested the tablets with 130 children who were out of school prior to the pandemic and had been for an average of nine months. Improvements were measured using localised EGRA and EGMA assessments, and results found that there were increases in the average percentages of correct answers across all subjects (6% literacy, 12% numeracy, 4% additional language). Another community (Lampung) tested the tablets with 456 children who were going to be tested in schools but were not currently attending school due to the COVID-19 closures. Again, results showed an increase in the percentages of correct answers across all subjects (5% literacy, 6% numeracy, 11% additional language), indicating that all children improved after using the app, but those who usually attended school made marginally less progress for literacy and numeracy outcomes, yet more progress for an additional language. These early results further reinforce the promise that tablet-based technology, coupled with learning software, has in improving foundational learning skills. However, the lack of formal analyses used to evaluate results makes it hard to assess the statistical significance of the findings (e.g. are there any significant differences between the improvements made between the out-of-school learners and the usually school-based learners?). Other factors important to interpreting the impact of an educational intervention were also missing, such as the magnitude of the effect and any variability in outcomes across the different learners (McGough & Faraone, 2009; Kraft, 2020).

Additional research suggests that tablet-based learning apps may be successful in improving learning outcomes. For example, 4.5 months use of the Mindspark learning app during the pandemic improved numeracy scores by 38% on average for 126,000 children in India (Mindspark, 2021; Gor, 2023). However, the implementation of such apps was dependent on the individual having access to their own devices, which is a previously documented concern for low-income children due to the cost and penetration of handheld technologies (Gor, 2023; Kukulska-Hulme, 2023; Rodriguez-Segura, 2022).

The potential benefits of remote learning for foundational skills during the pandemic are clearly evidenced but results seem to be inconsistent and dependent on the type of learners, context, and circumstances. These mixed results emphasise the need for conducting mixed-methods research to obtain more nuanced data that will help identify exactly which features of app-based learning are most successful in facilitating learning improvements (Nicolai et al., 2023). However, evidence gained from the COVID-19 pandemic is somewhat problematic as an experimental period because many other contextual factors changed at the same time that would likely contribute to learning effects, for example, the immense upheaval that the pandemic had on societal health, routines and economies (Bacher-Hicks & Goodman, 2021). It is also essential to note that emergency remote learning implemented with little notice differs from education designed to be delivered remotely (Crompton et al., 2021c). Findings from the remote provision during the pandemic have shown that various online mediums were used to facilitate lesson delivery, such as Zoom, Skype and Google Classroom, so that teachers could maintain routine contact and support where possible (Crompton et al., 2021b). In these circumstances, a balanced approach could be implemented in which technology and traditional learning methods were combined, as providing non-digital resources is an essential element of emergency remote education (Crompton et al., 2021b). Schools that have the necessary resources to provide instructional diversity were mailing out workbooks, involving parents, creating home-based activities, and incorporating physical activities to keep students engaged and provide much-needed structure from a distance, which are factors previously shown to facilitate learning attainment (Fairlie & Loyalka, 2020; Crompton et al., 2021b). The characteristics and challenges faced by children living in rural areas with limited access to quality schooling are distinctly different from those temporarily out of school due to the pandemic (Adukia & Evans, 2023; Giraldo et al., 2021). Children who have never attended or are consistently out of school face unique challenges, as discussed earlier in this chapter. Therefore, while COVID-19 research is insightful as to how EdTech may support remote learning, it is pertinent to build on these studies by specifically examining the impact of tablet-based EdTech interventions within a long-term, low-income, out-of-school population.

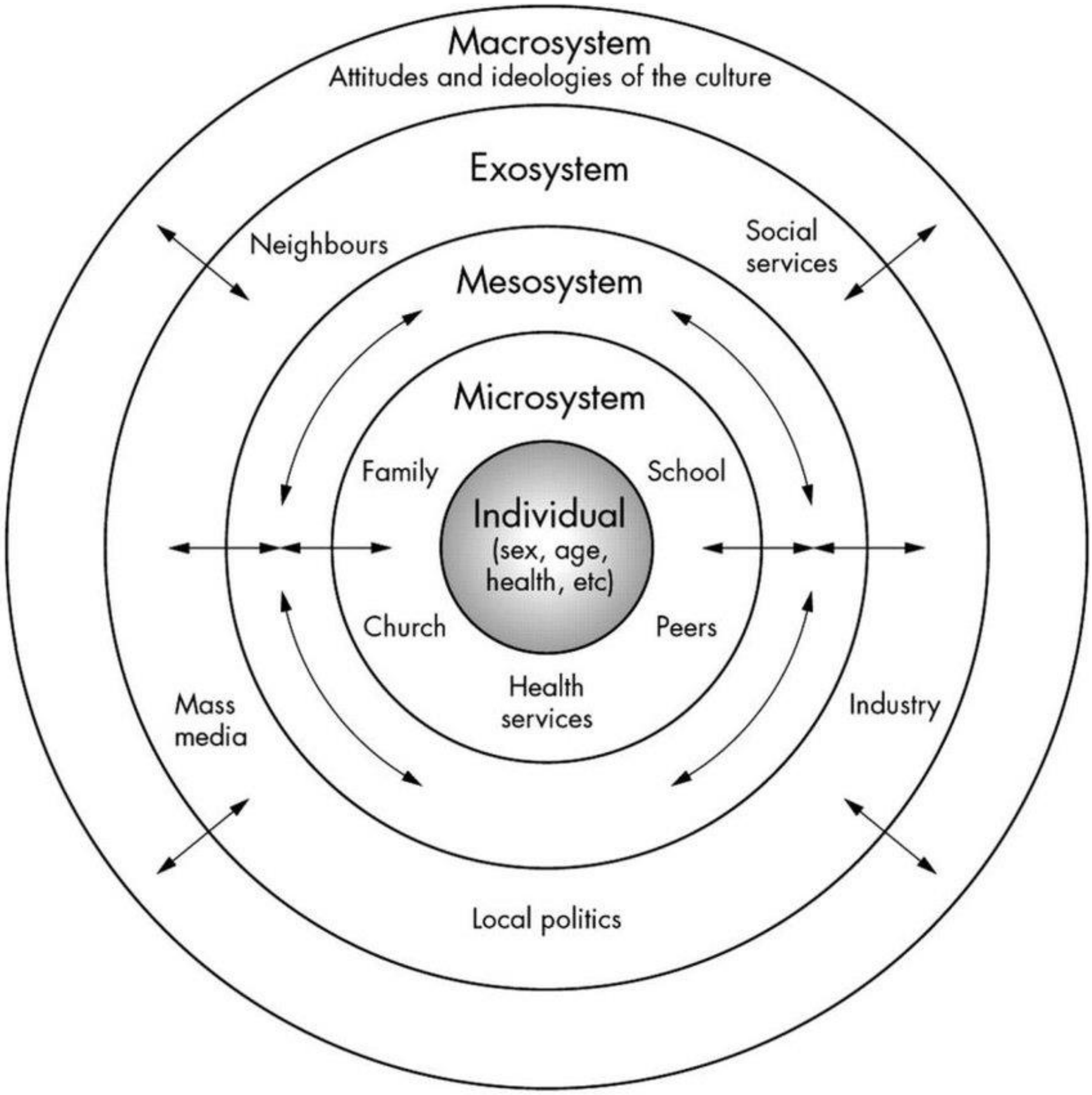
Theoretical framework

A multi-level model of interpretation is needed to provide a holistic explanation of EdTech learning interventions with out-of-school children, identifying which factors may influence children's foundational learning gains following an app-based intervention. A determinant framework could provide a strong foundation for developing an appropriately holistic model. This would synthesise research results by identifying barriers and facilitators to describe multi-level factors believed to influence implementation outcomes (Nilson & Bernhardsson, 2019).

One prominent theory is Bronfenbrenner's Ecological Systems Theory (1979; see Figure 1), which states that five interconnected ecological systems influence a child's developmental process. This ranges from their immediate surroundings to societal and cultural structures and is organised by how influential the setting is for the child. These five systems are the microsystem (e.g. family, school), mesosystem (connection of the microsystems), exosystem (e.g. local governments), macrosystem (e.g. politics, culture) and chronosystem (development over time; Bronfenbrenner, 1986; Edwards et al., 2016).

Figure 1

Bronfenbrenner's Ecological Systems Theory

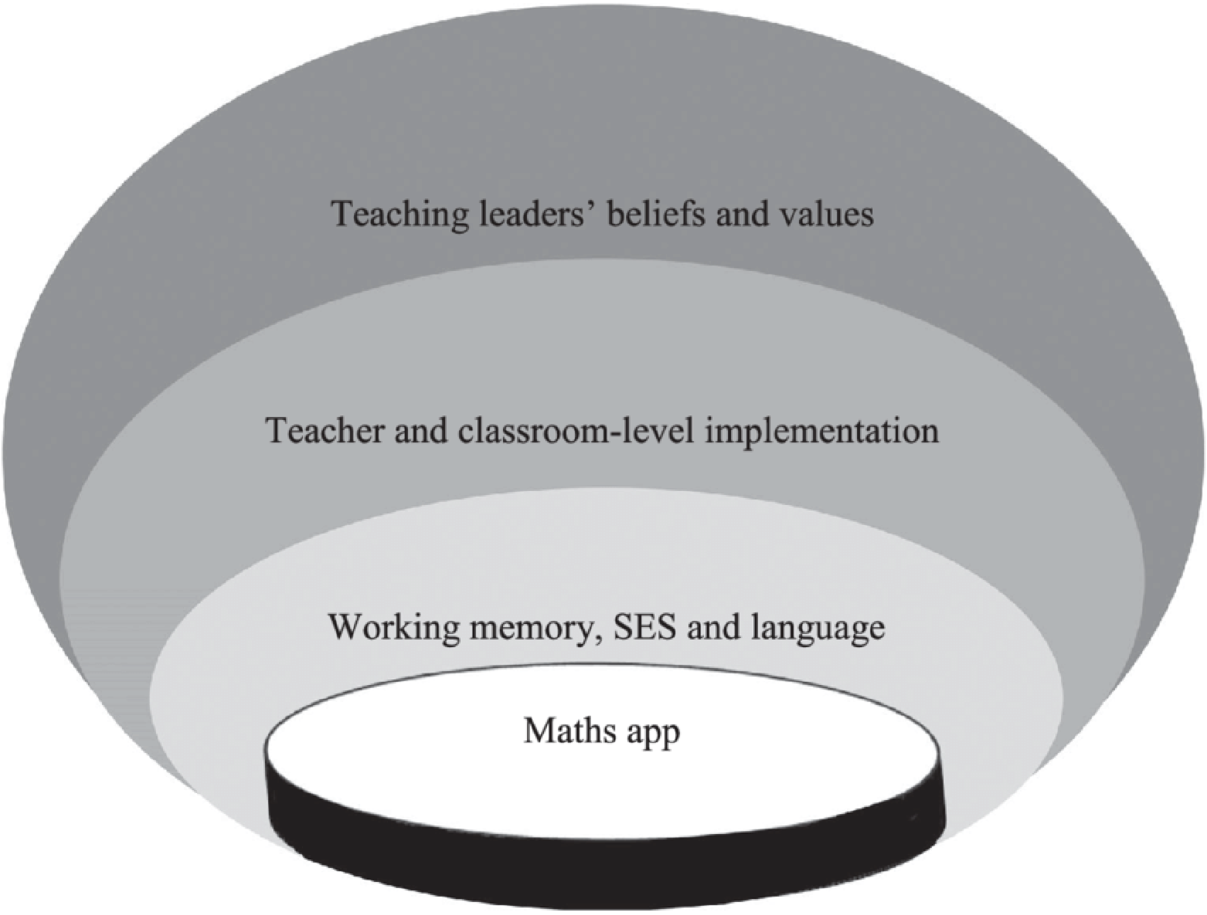


This theory has been an influential framework within developmental health research. It has also been adopted to help explain children's educational development, albeit in a traditional school setting (Eriksson et al., 2018; Lippard et al., 2017). It is commended as a practical, holistic approach to reflect the dynamic nature of complex, multi-level, relationships (Lopez et al., 2021). However, the ecological systems theory has faced some critique for being reductionist, as it minimises the individual agency of the child in the developmental process and places the child as a passive recipient with surrounding environmental and contextual influences that shape learning (Hammond, 2019). This perspective could limit the extent to which we can understand if, how, and to what extent children learn using technology in out-of-school settings. Therefore, the theory must be adapted to emphasise the out-of-school child and their unique characteristics, the cultural context, and any relevant elements of a digital learning environment for it to be a sound framework to holistically evaluate the implementation of an app-based intervention.

Outhwaite et al. (2019b) adapted the ecological systems theory in the context of an app-based maths intervention, adjusting the 'centre' of the model to be the learning outcomes following the intervention and adding individual child-level factors (e.g. SES, working memory; see Figure 2) to be the surrounding microsystems. This model was more focused on the context of app-based learning than Bronfenbrenner's original model, with the macro-level factors relating to teaching leaders' beliefs and values rather than examining app-based learning at the broader policy and cultural level (Outhwaite et al., 2019b; Outhwaite, 2019). Whilst Outhwaite's model placed app design and content as influential within the framework, they used two learning apps to test their model, the onebillion maths apps (Maths 3-5 and Maths 4-6), which had been previously shown to support children's learning (e.g. Outhwaite et al., 2017). Their research scope did not include identifying and evaluating individual app features that may influence learning outcomes. Additional exploration of different apps would be highly beneficial in informing app design by testing whether specific app features are more influential than others in facilitating learning improvements.

Figure 2

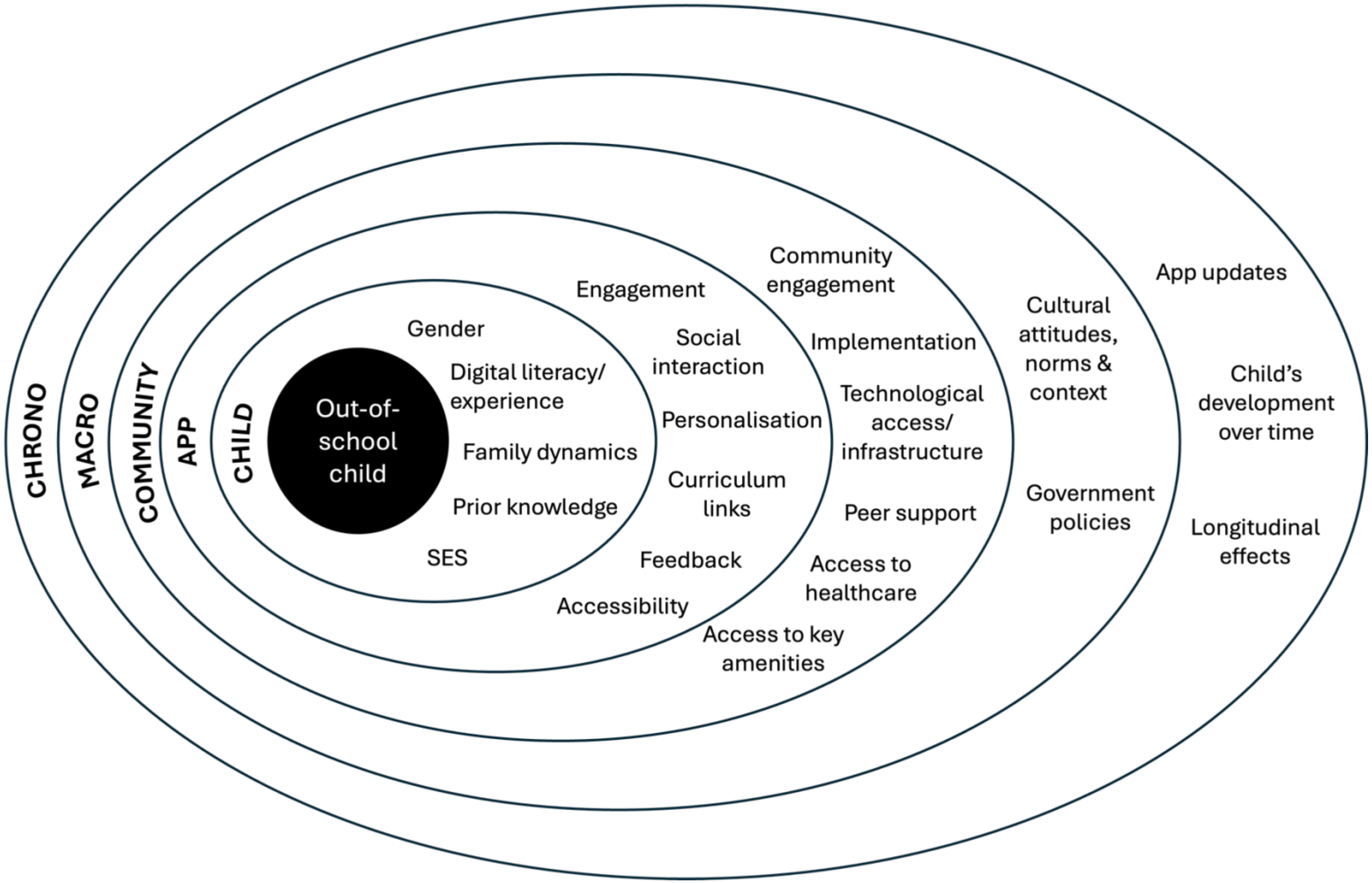
Outhwaite's multi-level model outlining factors that may influence learning outcomes following a maths app intervention



Furthermore, out-of-school children are a distinct group of learners with unique challenges that are not captured in existing models. Existing models needed to be modified and built upon to create a new model specifically for this population, as shown in Figure 3. This thesis utilised a mixed-methodological approach to develop an evidence-informed model that evaluates learning apps deployed in an out-of-school context at different levels of explanation. Particular focus was given to app features, child and community-level factors, and implementation. This framework allowed for a holistic oversight of implementing a learning app to help determine which interconnected factors may predict children's foundational learning gains when using tablet-based technology and learning apps in a marginalised community lacking resources and formal learning support.

Figure 3

A holistic multi-level model depicting the potential predictors of learning improvement for out-of-school children using an EdTech intervention



Thesis Aims

To address the research gaps identified above, this thesis evaluated the deployment of five learning apps in an out-of-school context in Tanzania, utilising the data, software, and insights gained from the GLXP competition (see Chapter 2 for further details). Throughout the thesis, potential predictors of foundational learning outcomes were examined following a tablet-based intervention deployed directly to the community.

Tanzania is a low-income country with over 1.4 million children and adolescents recorded as being out of school, with the true figure likely to be much higher (Jordan et al., 2021). In 2016, the Tanzanian government introduced a fee-free education policy for both primary and secondary schools, which decreased the number of out-of-school children. By 2019, primary enrolment was as high as 98.8% across the country (Jordan et al., 2021). The high uptake is unsurprising, considering that parental inability to pay school fees has been cited as one of the greatest barriers to educational opportunities in sub-Saharan Africa, particularly for girls (Ombati & Ombati, 2017; Lyanga & Chen, 2020).

Despite this, in 2021, secondary enrolment was only 32%, suggesting a high level of dropout between primary and secondary schooling, even with the significant barrier of school tuition fees removed (Jordan et al, 2021; Global Partnership for Education, 2023). Furthermore, providing school access and improving learning outcomes is still proving a challenge in Tanzania, particularly since the global COVID-19 pandemic (Global Partnership for Education, 2023; Joseph & Irhene, 2021). This level of dropout is worrying and suggests that children are either not engaging with the education offered in their local schools, or that contextual factors may be influencing children to drop out from school after an initial period of learning (e.g. competing household responsibilities).

However, at the end of the GLXP competition, 276 of the out-of-school children reported attending school at the time of the endline assessment, suggesting that autonomous learning with tablet-based technology may encourage children and their families to reconsider school attendance and be more interested in developing their foundational skills. As research has previously indicated, EdTech alone may enhance engagement and motivation to learn with out-of-school children, so even if attending school is not feasible, EdTech solutions can still provide children with an engaging, flexible method of learning (Stubbé et al., 2016; Afoakwah et al., 2021). Providing out-of-school children with a tablet-based learning solution, as the GLXP competition did, could serve as a pivotal first step to facilitating the children's entry into formal schooling. Where that is not possible, it could also offer the alternative of a well-designed learning app to bridge the attainment gap in the absence of traditional school settings. Therefore, this thesis research investigating the predictors of out-of-school learning was crucial to help provide evidence-informed recommendations for the future development of learning apps and successful implementation with this demographic.

The specific aim of this thesis was to identify potential predictors of foundational learning outcomes following a tablet-based intervention directly deployed to an out-of-school community in Tanzania. These predictors will be explored based on different levels of explanation, focusing on app features, gender, child-level, and community-level features, as outlined in Figure 3. This aim was addressed in Chapters 3-6. Chapter 2 provides an overview of the GLXP and the data that was collected and analysed in this thesis.

Chapter 2: Methodology

This thesis explores and evaluates the potential factors that might influence foundational learning with handheld tablets and learning apps for young out-of-school children based in rural Tanzanian villages over 15 months. The focus of this research is based on the case study of the GLXP competition, in which an educational technology intervention was implemented to assess the effectiveness of five learning apps to test their success at improving foundational reading, writing and maths skills (XPRIZE, 2019). This chapter focuses on the methodological background and case study for this thesis, starting with the chosen philosophical approach and justification for the methodological approach. This is followed by a detailed overview of the competition, the five finalist learning apps, and the resulting dataset collected during the GLXP competition that forms the basis of this research.

Philosophical position of the research

This thesis adopted an inductive mixed-methods approach, with methodological pragmatism as its guiding philosophical foundation, to investigate the multi-faceted phenomenon of out-of-school learning in Tanzania using educational apps delivered on tablet-based technology.

Methodological pragmatism is a pluralist methodological paradigm that offers a flexible approach to research design due to the overarching principle that no methods are intrinsically good or bad; instead, some methods are better suited to each specific research question and the needs of the research (Foster, 2024; Rescher, 1977; Teddlie & Tashakkori, 2009). This flexible, problem-solving approach to research design makes it highly suitable for complex, real-world applications like those central to educational research, as the epistemological underpinnings of methodological pragmatism are the value of knowledge for its practical implications and use in solving real-world problems (Foster, 2024; Rescher, 2017).

Methodological pragmatism is a uniquely balanced and open-minded position in which hypotheses developed from both positivist and critical perspectives can be integrated and processed holistically, with reflexivity, to provide multi-faceted explanations (Houston, 2014; Foster, 2024). This pragmatic epistemological stance does not automatically identify mixed methods as a superior approach but supports its use to generate a comprehensive understanding of how educational apps may facilitate learning in a marginalised, out-of-school setting (Foster, 2024; Clarke, 2021; Johnson & Onwuegbuzie, 2004). Integrating multiple levels of analysis, including conducting experimental primary research with inferential statistics, using machine learning methods for secondary data analysis, and qualitative interview techniques, allows for a nuanced exploration of the key factors influencing learning outcomes.

This methodological pluralism is a deliberate choice to embrace the complexity of the research topic and to ensure that the findings are robust, have relevant implications, and are realistically grounded in the learner experiences (Hesse-Biber, 2015; Teddlie & Tashakkori, 2009; Shannon-Baker, 2016). In the context of this thesis, it is a suitable choice to produce actionable knowledge that focuses on the practical application of the research findings, aiming to inform the development of effective EdTech for out-of-school children and guide policy and practice for digital learning solutions, ultimately contributing to the reduction of educational disparities.

Justification of methodological approach

Most EdTech intervention research has utilised randomised control trials (RCTs) in recent years as the central focus of funding and engagement for developing countries to address the attainment gap; approximately 50 percent of private sector funding in Syria has, for example, involved the development of technological education innovation (Menashy & Zakharia, 2017). Frequently, RCTs have been employed as pilots, a first step in attempts to secure further funding (Feeley et al., 2009) - and are often considered a 'gold standard' for evidence-based research and practice, given the need for rigorous research design in large-scale intervention implementations (Kratochwill & Levin, 2014). Furthermore, conducting RCTs as field experiments allows researchers to investigate the effect of an educational app in the context of its natural environment and, therefore, make appropriate changes and adaptations necessary for further real-world application (Allen, 2017).

However, while RCTs are valuable, they also present significant limitations, particularly when considering low- to middle-income countries. A major challenge is that these trials often overlook crucial contextual factors, such as local cultural attitudes towards technology, infrastructural limitations, and varying levels of digital literacy among participants (Tauson & Stannard, 2018; Buchanan et al., 2007). These factors can significantly affect the outcomes yet are rarely accounted for in the design of RCTs, leading to results that may not accurately reflect the conditions under which interventions will be implemented. As Burns (2022) points out, the effectiveness of EdTech can be highly context-dependent, and without incorporating these nuances, RCTs may yield findings that are difficult to generalise across different low-income settings.

Another significant barrier when conducting RCTs in low- to middle-income countries is not only the expense of logistics for large community-based interventions but the lack of guarantee that results will be significant (Buchanan et al., 2007) - a particular challenge to an underfunded area still requiring attention at a global scale. Whilst the effectiveness of EdTech interventions has been established through RCTs, some results have shown that positive gains often start to diminish over time, although there are early signs of sustained success in some school-based research (Tauson & Stannard, 2018; Outhwaite et al., 2017). Furthermore, using RCTs in isolation limits the insights gained from an intervention. While RCTs can establish whether an intervention works, they often fail to explain how and why it works, especially in educational contexts where various external factors may interact. For example, RCTs alone may not provide sufficient information on how to replicate an intervention in diverse contexts or account for factors such as teacher roles or community dynamics (Williams, 2020; Moore et al., 2015). This is particularly pertinent in low- and middle-income countries where conditions are variable and often unique, making it challenging to rely solely on quantitative data from RCTs (Nicolat et al., 2023). Therefore, additional research is needed to identify predictors of learning, address factors influencing the sustainability of effects in EdTech interventions, and to employ follow-up or longitudinal methods for interventions that allow long-term outcomes to be tracked effectively (Walton, 2018).

To gain a more comprehensive understanding of EdTech efficacy, it is essential to explore the experiences and perspectives of the individuals implementing these interventions. Qualitative methods, such as interviews and focus groups, are invaluable tools for assessing the acceptability and adaptability of interventions within specific target settings (Ayala & Elder, 2011). These methods allow for a deeper exploration of the role of the teacher and other local stakeholders in EdTech through interviews and focus groups (Clarke & Braun, 2014), isolating implementation challenges on the ground and enhancing teachers' strengths and roles within the learning environment. Furthermore, post-intervention research studies can inform long-term implementation and sustainability and provide insights into the general acceptability of the RCT within the target audience (Pegrum et al., 2013).

While there is a consensus that EdTech could contribute to foundational learning outcomes in low- and middle-income countries (Tauson & Stannard, 2018; World Bank, 2018), less is known about what drives the effectiveness of such technological approaches, leading to many initiatives being both designed and implemented without taking current evidence and local contextual factors into consideration (Tauson & Stannard, 2018). This gap is exacerbated by the lack of qualitative studies in these settings, making it difficult for investors to see the ‘bigger picture’ when deciding whether to choose and implement an intervention. Furthermore, a lack of randomised trials to support a given technology does not mean that a particular technology or software is not practical; there may be observational evidence or qualitative research demonstrating that the technology seems to be successful in a particular context, or a hard-to-reach setting may make it challenging to have obtained sufficient evidence thus far for diverse groups of learners (Burns, 2022; Zubairi et al., 2021).

Robust evaluations are needed to inform and change decision-making around implementing educational technology to support foundational learning, particularly in low-income countries (Major et al., 2021; Nicolai et al., 2023). The current most common method of establishing whether an EdTech initiative is effective is to measure the learning gains made during the intervention (Rodriguez-Segura, 2022). However, this evaluation method can be problematic due to the difficulty of measuring learning; only using test scores can be inadequate due to the lack of ability to determine cause-effect relationships (Spaull & Taylor, 2015; Burns, 2022). It is increasingly difficult for technological initiatives to isolate the driver of the learning gains, as the convergence of technologies means that when the whole initiative is novel, it is hard to identify whether the improvements are due to the hardware, software, or context (Burns, 2022). Furthermore, while measuring learning outcomes is crucial for evaluating the efficacy, this will not allow insights into the specific pedagogies, app features and contextual factors that may support the learning gains found, which limits the implications and conclusions that can be drawn from the findings (Pawson et al., 2005; Nicolai et al., 2023).

Ultimately, the research on impact evaluation demonstrates how EdTech initiatives are currently evaluated with learning outcomes and highlights the need to refine and develop evaluation methods to provide gold-standard evidence-based research on EdTech implementation. It has been demonstrated that while RCTs can offer useful information about the success of learning interventions, there are limits to the knowledge that can be gained, particularly around the impact that contextual factors and lived experiences may have on the learning process (Corr et al., 2018). Therefore, a mixed-method design would allow the selection and combination of the most appropriate qualitative and quantitative research methods to answer the research questions and provide a holistic evidence-based evaluation of the intervention (Fàbrugues et al., 2023).

As the use of EdTech increases rapidly to help children learn in low-income countries, there is increased potential for reliable data collection and more robust analysis using machine learning methodologies, which can then inform data-driven, effective decision-making and improve educational programmes at local and national levels (Longley, 2022; EdTech Hub, 2022). Previously, the poor availability and use of data have negatively impacted decision-making and the aim to improve educational outcomes in low-income countries (Longley, 2022). As a result, the EdTech Hub has made data-based decision-making an explicit focus in its current work due to the hugely beneficial impact it is anticipated to have on the quality of educational instruction for countries with limited resources (Longley, 2022; EdTech Hub, 2022). However, a pivotal criticism of big data research is that it is somewhat limited to correlational models and predictive analytics, which makes it difficult to determine causality for educational research and explain the behaviour behind the predictions (Daniel, 2019; Lavelle-Hill, 2020).

As a result, this thesis employed a multi-methodological approach to address the key limitations of each research method, utilising a novel approach that combines experimental methods, inferential statistics, machine learning methods, and qualitative research. While RCTs and machine learning techniques can be used to assess learning gains and identify specific app features and contextual factors that predict outcomes, qualitative research provides nuanced insights into the lived experience of participants and the contextual variables that quantitative methods alone cannot capture. Learning gains and accompanying data will be used to isolate the app features (Chapters 4 and 5), contextual (Chapter 6), and implementation factors (Chapters 6 and 7) that may drive learning outcomes, using the GLXP competition as a case study.

This thesis used machine learning methods to determine the app, child and village-level features that predict learning improvements in out-of-school children. However, the mixed methods approach aimed to address the identified criticism of machine learning approaches by introducing qualitative research to explore and determine some of the explanations behind why and how children learnt within the EdTech intervention and how implementational barriers can be addressed in future EdTech research with this demographic. The findings from this mixed-methods approach can guide the design and implementation of EdTech interventions that are both evidence-based and context-sensitive, facilitating the development of more sustainable and impactful EdTech (Fàbrugues et al., 2023; Corr et al., 2018).

Global Learning XPRIZE competition: Case Study

To address the global learning crisis described in Chapter 1, the XPRIZE Foundation - a non-profit organisation that designs incentivised competitions in areas such as space, education, and healthcare to try and bring people closer to an improved and more sustainable world - launched a ‘Global Learning’ competition in 2014 in partnership with the World Food Programme (WFP) and UNESCO. The competition challenged multi-disciplinary teams worldwide to develop open-source, scalable learning software to empower children to teach themselves basic literacy, numeracy and writing skills (XPRIZE, 2019). The challenge was to develop a tablet-based, scalable, self-guided digital technology solution for marginalised out-of-school children. XPRIZE placed no requirements or constraints on app structure or content, allowing the teams to be independent in their design.

The competition launch led to 198 teams from 40 countries registering to solve the challenge; all of whom were given 18 months to develop software before submitting their attempts at a learning solution to an expert international judging panel (XPRIZE, 2019). Eleven semi-finalists were selected to present a live demonstration to the judges and were evaluated on their creation and implementation of an engaging and autonomous design, ability to measurably increase the children's learning skills and creation of open-source, scalable and replicable software. Five finalist teams were selected to field test their app software with 2500 illiterate children aged 7-11 years, most of whom had never attended school, from across 172 remote villages in the region of Tanga, Tanzania, over 15 months (XPRIZE, 2019). Each village received one of the five finalist learning apps on hand-held tablets. Villages were allocated as equally as possible across the five finalist teams, so each team worked with 30 villages. There were also 22 control villages which received no tablet intervention during the competition. Each of the intervention villages received the software only from their dedicated team, and allocations were made to assure a statistical balance of distribution, where possible, amongst age, gender, and proximity to the solar charging stations that XPRIZE installed to enable the tablets to be charged throughout the field trial and beyond.

Throughout the competition, there was very little direct involvement from XPRIZE and the finalist teams to encourage autonomous learning with the software. Within each village, there was a nominated village 'Mama' or 'Baba' who were paid a stipend to ensure children had the correct equipment to learn, including ensuring the solar station was working and maintained, children had working tablets, and no one stole the tablets (Huntington et al., 2023a). They were required not to direct or assist the learning process. The children were not given a schedule for when or how often they should interact with the technology; instead, they were left to direct and organise their own learning autonomously. The competition employed local drivers to collect usage data from the charging points weekly, which was sent to XPRIZE. They also informed UNESCO and WFP of any technical support needed from the villages.

An independent evaluation conducted by RTI International assessed the impact of each app provided by the five finalist teams on foundational learning outcomes (King et al., 2019). Using standardised instruments (Early Grade Reading Assessment - EGRA and Early Grade Mathematics Assessment - EGMA), foundational literacy, numeracy, and writing skills of children participating in the field trial were assessed before implementation and again directly after the 15-month test period. Results indicated that solutions produced by KitKit School and onebillion achieved the greatest overall proficiency gains - although all teams demonstrated significant core improvements. Prior to the field test, 74% of the tested children never attended school, and 90% could not read any Swahili; these figures were halved at the end of the testing phase (XPRIZE, 2019). UNESCO also conducted a social-emotional assessment pre- and post-competition and found that, compared to controls, children who interacted with the software solutions showed improvements in many areas, such as self-esteem, self-expression, confidence, and independence (Shukia et al., 2019).

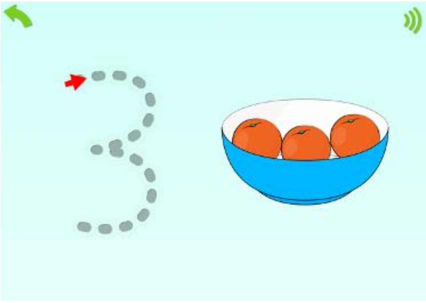
Whilst the GLXP competition provided evidence that interactive apps delivered through hand-held tablets could foster improvements in foundational learning for out-of-school children in remote regions of Tanzania, little is known about how this educational technology can be effective in this context. Understanding factors associated with positive learning outcomes in remote settings with out-of-school children is vital to enhancing theoretical understanding and promoting effective interventions. Accordingly, this thesis aims to use the full dataset provided by the XPRIZE Foundation in combination with additionally sourced new data to comprehensively explore the app-based, child-based, contextual and implementation factors that influence foundational learning outcomes by out-of-school in marginalised communities in Tanzania.

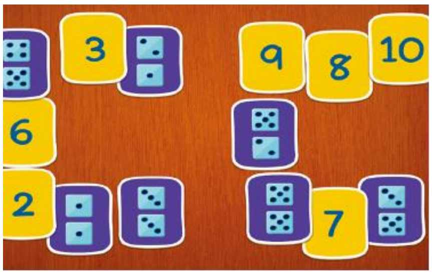
The learning apps

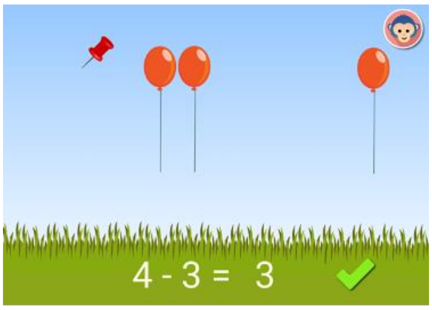
A brief description of each of the five finalist learning apps is outlined in Table 2. The learning apps were delivered on Google Pixel C tablets with a 10.2” screen with a 2560 x 1800 pixels resolution. Some finalist teams were developing learning tools before the competition (e.g. Enuma, who created KitKit School), and others were established specifically to address the XPRIZE call for action (e.g. Robotutor; XPRIZE, 2017).

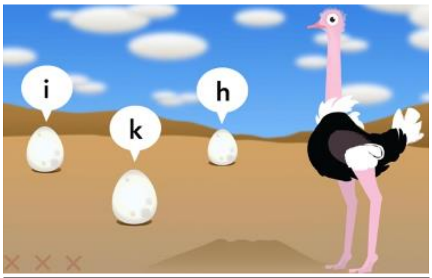
Table 2

Brief description of the five finalist learning apps from the GLXP as described on the organisations' websites (retrieved October 2022)

App/Website	Pedagogical Description
onebillion www.onebillion.org 	The child works through an ordered set of learning units covering reading, writing and numeracy. Each is a short interactive activity to teach or provide practice for a particular concept or skill. The app provides each child with an individual journey. Regular tests enable the app to deliver the best course through the content for each individual child based on their learning level. **Core skills:** Literacy – letter recognition, phonics, word construction, sentence formation. Numeracy – counting, number recognition, basic arithmetic, problem-solving. **Example learning activities:** Tracing letters and numbers Counting objects with visual aids Matching letters to sounds Basic phonics exercises Adding and subtracting with visual cues Regular quizzes to adjust learning paths

App/Website	Pedagogical Description
Kitkit School www.kitkitschool.com 	A sequenced progression of core literacy skills, from letter recognition to phonics and print awareness. Sequential courses introduce new skills and reinforce previously covered concepts at more difficult levels. Learning is scaffolded to support cognitive development and independent learning and accessibility functions engage and empower the world's diverse learners. **Core skills:** Literacy – alphabet recognition, phonics, reading comprehension, vocabulary building. Numeracy – number sense, counting, basic arithmetic, geometry, spatial reasoning. **Example learning activities:** Tracing letters Recognising letter sounds Forming simple words Reading stories Reading comprehension questions Basic counting exercises Addition and subtraction problems Visual puzzles/games for shape recognition and pattern formation

App/Website	Pedagogical Description
Chimple www.chimple.org 	Chimple features a series of educational games in literacy, maths, and digital skills. The app is easy to use and understand. It enables children to learn, practice and improve their reading, writing and math skills in a fun and interactive way. **Core skills:** Literacy – Letter formation, phonemic awareness, early reading skills. Numeracy – counting, patterns, arithmetic, logical thinking. Digital literacy – basic navigation, problem-solving through digital interfaces. **Example learning activities:** Letter tracing Phonics matching games – sound to letters Matching words to images Counting objects Recognising number sequences Basic arithmetic problems Logic-based games/puzzles Trial-and-error digital tasks

App/Website	Pedagogical Description
CCI www.cciny.net  	SchoolHouse (by CCI) includes a series of structured and sequential instructional lessons, as well as a platform that enables non-coders to develop engaging learning content in any language or subject area. **Core skills:** **Literacy –** Alphabet recognition, phonics, sentence building, early reading skills. **Numeracy** – Counting, arithmetic, problem-solving, basic geometry. **Example learning activities:** Sound-to-letter phonics games Counting objects and basic addition/subtraction Drawing exercises to develop fine motor skills Shape recognition Sentence building activities Gamified challenges Enhanced profile customisation for engagement (pictures, drawings etc.)

App/Website	Pedagogical Description
Robotutor www.cmu.edu/scs 	Robotutor is an open-source Android tablet app that enables children ages 7-10 with little or no access to schools to learn basic reading, writing and arithmetic without adult assistance. **Core skills:** **Literacy –** Letter recognition, phonics, vocabulary development, reading comprehension, sentence construction. **Numeracy –** Counting, basic arithmetic, fractions, geometry, problem-solving. **Learning activities:** Phonics matching exercises – sound to letter Reading comprehension – short stories and multiple choice questions Sentence construction writing activities Counting games Arithmetic problems Working with fractions and geometric shapes Gamified challenges to motivate learning (e.g. earning rewards)

The main dataset

Overview

This section introduces the main dataset that was generated during the GLXP and made available to the research team by XPRIZE as part of the collaborative agreement for this studentship.

The data includes detailed demographic and contextual data of individual participants and the full answers for the main proficiency tests - Early Grade Reading Assessment and Early Grade Math Assessment - at baseline and endline. Baseline data was collected in August 2017, and endline data was collected in March 2019. Data for 2041 children for whom a complete set of baseline and endline tests was available was investigated out of the ~2500 children initially recruited to participate in the GLXP competition (XPRIZE, 2019).

Following the competition, the full dataset was made available online for XPRIZE teams to peruse, but the data was anonymised so that it was not clear which app the children received. Teams also had access to a non-anonymised version of the dataset for optional independent analysis, with only 75% of the data points included. These anonymised datasets were made available through an online XPRIZE community forum that anyone could access if they created an account, although this resource has since been taken down from the XPRIZE community forum website. Only the research team associated with this studentship had access to the full main dataset, as well as the research team at RTI who conducted the impact evaluation of the field trial. The data provided for the studentship does not include the socio-emotional data that UNESCO collected throughout the competition process.

The app software that was used for the GLXP is open-source and available for download in the XPRIZE GitHub repositories (XPRIZE, 2020). The full tablet usage data for each team's villages can also be found in the repositories, but this was not utilised in the current research.

Literacy and numeracy data

The Kiswahili version of the Early Grade Reading Assessment (EGRA; Brombacher et al., 2014) and Early Grade Math Assessment (EGMA; Brombacher et al., 2014) were the psychometric instruments administered to measure foundational literacy and numeracy skills of individual children participating in the GLXP at both baseline and endline of the competition. These are standardised assessments commonly used in international comparison research for children between 5 and 15 years. RTI International used these assessments to measure learning outcomes for participating children to compare the effectiveness of five finalist apps to controls who did not receive an app-based intervention.

EGRA included subtasks of print awareness (out of 3), syllable sounds (out of 100), familiar word reading (out of 50), invented word reading (out of 50), oral reading fluency (out of 42), reading comprehension (out of 5), listening comprehension (out of 5), and a dictation-based writing task (out of 6). The total possible raw score was 261. EGMA included subtasks of number identification (out of 20), number discrimination (out of 10), completing patterns of missing numbers (out of 10), word problems (out of 7), addition (including three Level 2 questions; out of 23), subtraction (including three Level 2 questions; out of 23), patterns (out of 4), and identifying shapes (out of 4). The total possible raw score was 101.

Tanzanian versions of EGRA and EGMA tests have been shown to have good internal consistency (Cronbach's $\alpha = 0.79$ for EGMA, 0.94 for EGRA (English), and 0.98 for EGRA (Kiswahili; Brombacher et al., 2014)).

EGRA and EGMA tests were administered to the children by a trained researcher in their preferred language (English/Kiswahili), with their caregiver present. Some subtasks were timed as instructed in the guidelines (e.g. syllable sounds, addition).

Contextual data

Before administering EGRA and EGMA, researchers conducted a contextual survey with each participating child and their caregiver at baseline and endline. The survey included questions about the child's demographics (e.g. child age, child gender), family (e.g. the number of siblings, caregiver's profession), environment (e.g. the number of children in the village, access to electricity and water, remoteness from a school or city centre) and previous experience with technology (e.g. caregiver's phone type, exposure to a tablet before the GLXP). Data was collected with assent from the participating child and consent from their family. Questions could be skipped if the child or caregiver did not know the answer or did not understand the question. The full list of contextual questions asked can be found in Appendix B. A list of each of the additional variables (the EGRA/EGMA test questions) that were reported in the main XPRIZE dataset is provided in Appendix C.

Survey data descriptives

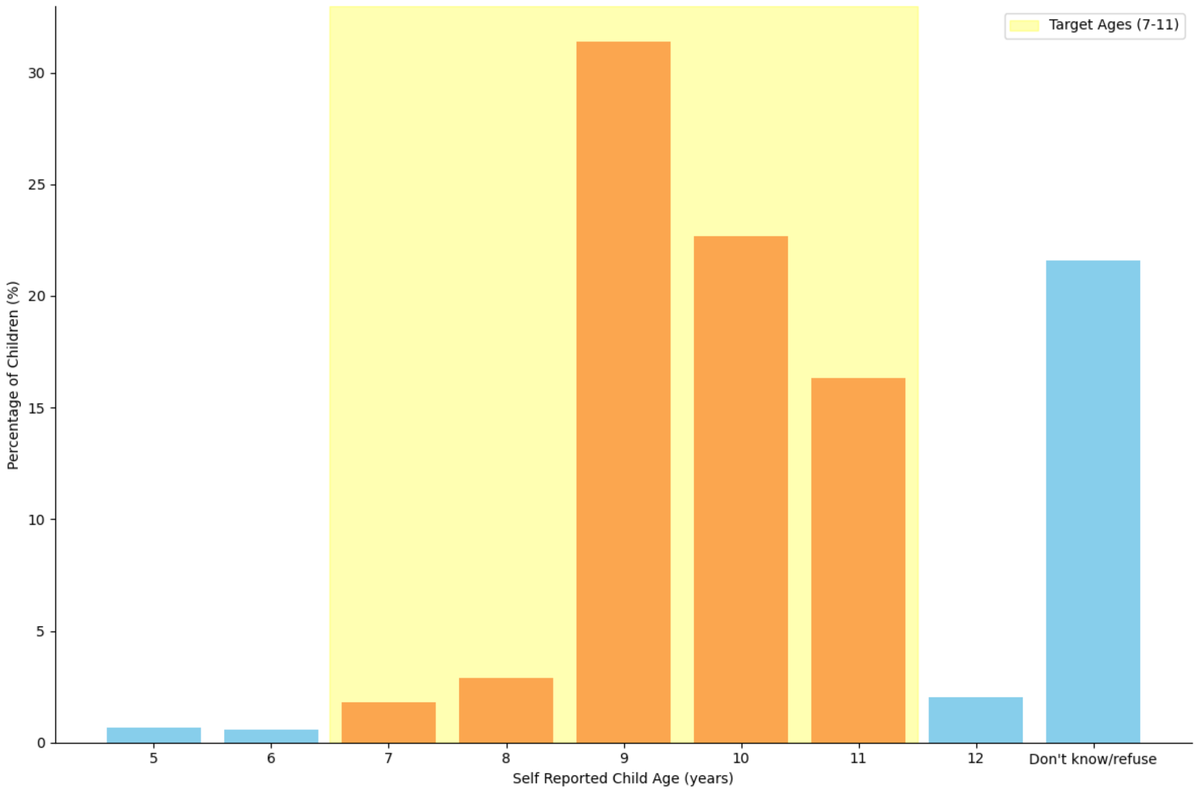
This thesis focused on key areas of interest from the contextual survey that the literature review (Chapter 1) suggested could be important for this population and setting. Chapters 5 and 6 also draw on this information where further information is required. Descriptives are provided below for these features, namely gender (number of girls and boys in the field trial), age (in years), children's experience with schooling (whether they have attended before, whether they attend now), technology (experience with smartphones and tablets), caregiver occupation and highest level of schooling, and the number of siblings (and whether they are educated).

Out of the 2041 children who completed the proficiency tests and accompanying survey, 762 were girls (~37%), and 1279 were boys (~63%).

The competition aimed to implement the learning intervention with children aged 7-11 years. Figure 4 shows the actual age distribution of the children tested (age given at baseline). As shown, 1.28% of children were below the target age, and 2% were higher than the target age. However, these ages are self-reported by the children, so are not necessarily 100% accurate. A large number of participants (441; 21.6%) either did not know their age or refused to answer.

Figure 4

Histogram showing the age distribution of participants at baseline



The questionnaire asked about the children's prior experiences with schools and technology. Many respondents reported that they had never attended school before (N = 1505; 74.74%), with 536 (25.26%) having previously experienced some form of schooling.

At the time of the endline experiment (March 2019), the children were asked if they were currently attending school. Many children responded that they were not (N = 1765; 86.48%), but 276 (13.52%) responded that they currently attended school. There is no data to suggest when they started attending school nor an indication of whether it had anything to do with being part of the GLXP competition.

The majority of children reported that they had never used a smartphone before (N = 1981; 97.06%), 52 (2.55%) had used a smartphone sometimes, and eight (0.39%) reported using them daily. When asked about handheld tablets, experience was even lower, with most never having used one (N = 2009; 98.43%), a few participants having used a handheld tablet sometimes (N = 28; 1.57%), and four respondents having used them every day prior to the competition.

Children were also asked about their home literacy experiences. Most children did not read aloud to anyone at home (N = 1806; 88.49%); 190 (9.31%) read aloud sometimes, and 45 (2.2%) read aloud to someone in their household daily. A similar pattern could be seen for the children being read to by another person, with 1710 children (83.78%) saying that it did not happen, 269 (13.18%) saying they were read to sometimes, and 62 children (3.04%) on a daily basis.

Some of the survey questions were directed at the caregiver who had attended the testing session with the child. The caregivers reported their occupation type as one of the following: “Unemployed/do not conduct work that generates income” (N = 120; 5.88%), “Agriculture, farming, livestock or fishing” (N = 1597; 78.25%), “Informal sales (sell food, crafts or other goods from home or market” (N = 134; 6.57%), “Construction worker” (N = 8; 0.39%), “Formal business owner” (N = 24; 1.18%), “Teacher” (N = 9; 0.44%), “Other professional job (NGO worker, manager, office assistant etc.; N = 116; 5.68%)”, “Traditional doctor or healer" (N = 11; 0.54%), or "Other" (N = 22; 1.08%). As the counts indicate, the vast majority of caregivers worked in an agricultural occupation.

The caregivers were also asked about the highest level of schooling they have reached. A large number of caregiver respondents had not completed any schooling (N = 736; 36.06%), the majority had reached a primary school grade (N = 1202; 58.89%), 78 (3.82%) had reached secondary school, three (0.15%) reached A Levels, five (0.24%) reached a tertiary vocational certificate, ten (0.49%) reached college or university, and seven (0.34%) responded "other".

Several questions were also asked about the children's siblings. The number of siblings a participating child had ranged from 0-20, with a mean of 3.7 (SD = 2.12). When asked how many of the children's younger siblings knew how to read and write, the majority said none (N = 1792; 87.8%), while 160 (7.84%) stated that some of the younger siblings knew how to read and write, and 89 (4.36%) claimed that all of the younger siblings knew how. For older siblings, 810 caregivers (39.69%) answered that none of the older siblings knew how to read and write, 412 (20.19%) answered that some knew how to read and write, and 819 (40.13%) said that all of the older siblings could read and write. Overall, these descriptives highlight the lack of access to and prior education that the sample demographic has experienced so far, emphasising the novelty and opportunity that a tablet-based learning intervention would have provided to these children.

Summary

This chapter has introduced the foundations for this novel mixed-methods thesis, including the case study on which the thesis is based. Table 3 summarises the key research questions addressed throughout this thesis, the chapter in which they are investigated, and the datasets and methodology chosen to address the question.

Throughout the following four empirical chapters, each chapter utilises a different methodology, the specific details of which can be found in the chapters' method sections. Further data was collected for Chapters 3, 5 and 6, which were used in combination with the GLXP dataset to gain a more comprehensive understanding of what factors drive learning improvements with autonomous app-based learning. The results from each chapter and different methodology will ultimately be triangulated to address the broader aims of the thesis, outlined in the introduction, and present a holistic model of the key app-based and contextual predictors, barriers and implementation strategies for out-of-school app-based learning in rural areas of Tanzania.

Table 3

Summary of research questions that are addressed in this thesis

Research questions	Thesis chapters	Areas of focus and dataset(s) used	Methods
Which features characterise each of the five finalist learning apps of the Global Learning XPRIZE competition (GLXP)? Do the five finalist learning apps of the GLXP that have been shown to support positive learning outcomes with out-of-school children within the same RCT share features?	Chapter 3 - Comparative Judgement of Pedagogical Features in Interactive Apps	App features *Dataset:* Primary dataset collected from experiment	Within-participants experimental design *Inferential statistics:* Binomial tests of probability Spearman's rank correlation

Research questions	Thesis chapters	Areas of focus and dataset(s) used	Methods
Which app features most predict improvements in literacy and numeracy outcomes for out-of-school girls and boys in Tanzania following an EdTech intervention? Are there significant differences in the impact of individual app features on literacy and numeracy outcomes for out-of-school girls and boys?	Chapter 4 - Relative Importance of App Features across Domains and Genders	App features Domain differences Gender differences *Datasets*: GLXP main dataset Dataset collected for Chapter 4	Secondary data analysis of XPRIZE dataset and Chapter 4 Machine learning analysis *Inferential statistics:* 2x2 ANOVAs and post-hoc t-tests

Research questions	**Thesis chapters**	**Areas of focus and dataset(s) used**	**Methods**
Which factors best predict foundational learning skills in out-of-school children in remote areas of Tanzania? Which factors best predict improvements in foundational learning outcomes after implementing an EdTech intervention with out-of-school children in remote areas of Tanzania?	Chapter 5 - Contextual Predictors of Foundational Learning Outcomes	Child-level predictors Contextual predictors *Datasets*: GLXP main dataset New contextual dataset	Secondary data analysis of XPRIZE dataset (baseline data) integrated with independently collected geospatial data from online sources. Machine learning analysis

Research questions	Thesis chapters	Areas of focus and dataset(s) used	Methods
How do experts associated with the GLXP perceive the impact of EdTech in supporting autonomous learning in remote low-income settings? What were the key challenges and opportunities identified by experts during the implementation of the GLXP competition?	Chapter 6 - Expert Elicitation of the GLXP competition	Implementation Context Lived experiences *Datasets:* New qualitative data collected for the chapter	Semi-structured interviews Thematic analysis

Chapter 3: Comparative Judgement of Pedagogical Features in Interactive Apps

Background

As outlined in Chapter 1, a growing evidence base suggests that using curriculum-based apps can provide high-quality education to children globally, supporting positive outcomes in literacy development, maths, science, problem-solving and self-efficacy (Bettinger et al., 2020; Herodotou, 2018). Literacy and numeracy interventions in developing and developed countries, implemented at school and in a home environment, have shown significant improvements in children's attainment, learning outcomes, and motivation (Major & Francis, 2020; Stubbé et al., 2016). However, no prior research has evaluated what app features may facilitate learning for out-of-school children, which makes it difficult to design effective apps for this population. This study utilises a comparative judgement task undertaken with 41 naïve participants who were asked to directly compare the five finalist apps of the GLXP on 15 pedagogical app features to determine what app features are most successful in facilitating foundational learning skills for out-of-school children.

Importance of designing effective learning apps

There was an exponential increase in mobile learning apps across a three-month period in 2019-2020, with over 900 million learning apps downloaded worldwide (Statista, 2021a). As the COVID-19 pandemic took a grip on educational provision globally, resulting in school closures that left approximately 1.2 billion children unable to attend school worldwide (Forbes, 2021), dependence on tablet-based apps increased further, and emphasised the need for alternative, effective education provision that is feasible when schools are closed, or children are not able to attend school (Azevedo et al., 2021).

When successful, interactive apps can offer exciting and effective learning environments that foster child-centred learning and starkly contrast traditional teacher-centred learning styles (Papadakis & Kalogiannakis, 2017; Ting, 2015). Learning apps currently on the market vary vastly in structure, content, and quality, so unless they are evaluated scientifically, researchers, educators, and parents cannot establish how effective apps are in supporting learning (Kolak et al., 2021).

Effective pedagogical features

Despite an accelerated uptake of mobile learning apps, relatively little is known about which app features support positive learning outcomes (Kim et al., 2021). When apps are used outside of the school setting, without the support of a teacher or caregiver to promote the acquisition of foundational skills, it is vital to decipher app features that are effective in assisting children's learning.

Over the past decade, several app features have been proposed to support learning, as mentioned in Chapter 1, derived from evaluation tools created to assess the design and potential of learning apps (e.g., Kolak et al., 2021; Papadakis et al., 2017; Outhwaite et al., 2023). Well-chosen interactive features embedded into mobile apps can help facilitate child-paced, inclusive learning environments. Table 4 describes 15 key app features that different researchers have attributed to support learning and contribute to the educational value of an app. Whilst not an exhaustive or partially systematic review of each paper, these features reflect a categorisation of attributes commonly identified in the literature as supporting learning until saturation for the purpose of this study was reached. Only pedagogical app features are reported; technical characteristics of tablets that may impact learning, such as screen size, are not considered.

Table 4

App features that have been attributed to support children's learning in the literature

App Feature	Definition	Cited By
Active Learning	Is the app 'minds on'? Does it require thinking or intellectual effort (rather than just cause-and-effect interactions or guessing)?	Hirsh-Pasek et al. (2015) Plump & LaRosa (2017)
Engagement in the Learning Process	Do the features engage you in app activities? Or do they distract you? Are the visual and sound effects excessive? Are there any disruptive ads?	Hirsh-Pasek et al. (2015) Plump & LaRosa (2017) Lee & Cherner (2015)
Meaningful Learning	Is the content meaningful and relevant to childrens' everyday experiences? Is it taught in a manner that can be contextualised within existing knowledge?	Hirsh-Pasek et al. (2015) Kolak et al. (2021)
Social Interaction	To what extent can the children interact meaningfully with: (a) characters through the app interface, and (b) caregivers around the app?	Hirsh-Pasek et al. (2015) Kolak et al. (2021) Lee & Cherner (2015)

App Feature	Definition	Cited By
		Outhwaite et al. (2022)
Accessibility – Language Demand	Is the language used simple enough to be accessible to children with Special Educational Needs and Disabilities (SEND) or children with lower language proficiency?	Kolak et al. (2021) Gulliford et al. (2021)
Accessibility – Motor Skills	Would the app be suitable for children with lower motor skills? Would it be good for children who usually struggle with traditional paper and pencil skills?	Allison (2019) Gulliford et al. (2021) Pitchford et al. (2016) Pitchford et al. (2018)
Accessibility – Autonomous Learning	Is the app easy to navigate independently and signposted appropriately for someone with limited tech experience? Is the avatar (if there is one) helpful in guiding the child? If no external caregiver, do you think a child could navigate the app with some trial and error?	Papadakis et al. (2017) Lan (2018) De Raynal et al. (2020)

App Feature	Definition	Cited By
Task Structure	Are the tasks structured in a way that makes sense? Is the child directed to which tasks they should complete next? Is there opportunity to reinforce previously learnt skills/knowledge?	Gulliford et al. (2021) Outhwaite et al. (2022) Callaghan & Reich (2018)
Task Processes – Feedback	Does the app provide both positive and negative feedback? When negative, does the error signal come with linked instructional feedback (to help them understand what they're doing wrong)? Is the feedback encouraging and potentially exciting for children (e.g., getting a prize for completing a task)?	Kolak et al. (2021) Gulliford et al. (2021) Outhwaite et al. (2022) Benton et al. (2021)
Curriculum Links	Do you think the tasks are close to what would be taught in schools on the curriculum? Do you think they're necessary topics/things children should know? Do you think the apps could be used to reinforce things taught in-class? Do you think the apps could be used as an assessment tool for curriculum modules?	Gulliford et al. (2021) Richards (2015)

App Feature	Definition	Cited By
Gamification	Are there games within the app? Are these fun and engaging? Is there a reward for 'winning' the game or being successful, or different levels to complete?	Putra et al. (2018) Lee & Loo (2021) Al-Azawi et al. (2016)
Personalisation/ Personalised Levelling	Is the app personalised to the child? Does it provide tasks that are specific to the child's level of learning? Do these tasks adjust accordingly as the child progresses? Or is it a 'one task fits all' approach?	Outhwaite et al. (2022) Lee & Cherner (2015) Benton et al. (2021) Vanbecelaere et al. (2020b) Vanbecelaere et al. (2021)
Retrieval-Based Learning (RBL)	Are there quizzes/tasks to test what children have previously learnt in the app?	Pitchford (2015) Grimaldi & Karpicke (2014)
Direct Instruction	Does a teacher/avatar give explicit instructions for how to do something? Do they demonstrate this or is there an example shown, using a demo, written example, or video? Do they use the 'show, try, test' method?	Outhwaite et al. (2019a) Toub et al. (2016) Chodura et al. (2015)

App Feature	Definition	Cited By
Free Play	Is the play unstructured? Are children free to explore as they want to? Are the activities child-centred? Do any of the tasks allow children to use their imagination?	Toub et al. (2016) Hirsh-Pasek et al. (2015)

Although educational researchers have championed these app features, it is not possible to know how important these features are in supporting learning. To explore this, direct comparisons of different apps that have been deployed in the same context with known learning outcomes need to be made. This is crucial, as some of the app features directly contradict each other. For example, direct instruction takes a prescribed, structured approach to learning, whereas free play is more child-centred and provides children with the autonomy to explore in an unstructured manner (Hirsh-Pasek et al., 2015; Outhwaite et al., 2019a). Playful learning that combines direct instruction with free play has been demonstrated in school environments and with app-based learning and could be highly effective in promoting socio-emotional and cognitive development in primary-aged children as both are shown to have unique benefits (Hirsh-Pasek et al., 2015; Toub et al., 2016). Apps that embody both features afford instruction through gradual release of responsibility, whereby cognitive work slowly and gradually shifts from tutor-led direction to student-led discovery (Fisher & Frey, 2021; Northrop & Killeen, 2013). Hence, it is important to identify the combinations of features embedded within mobile apps that have been shown to be effective at promoting learning outcomes.

To understand how individual or combinations of app features support learning, app features need to be linked directly to learning outcomes. A first step towards this goal was made by Outhwaite et al. (2022) in a systematic review and content analysis of maths apps targeted at the first three years of compulsory schooling that had been previously evaluated in the literature, as well as the Top 25 learning apps on the App Stores. They conducted a qualitative comparative analysis (QCA), which identified specific app design features and combinations of features that were shown to be sufficient to support children's learning of maths with educational apps. Out of 50 studies included in their systematic review, only 8 apps met the criteria required for the QCA of having data reported in sufficient detail to enable within-subject effect sizes (Cohen's d) to be calculated. Results of the QCA revealed that features of the maths apps that promoted programmatic levelling, such as scaffolding and personalisation, as well as explanatory and motivational feedback, maximised children's learning outcomes. Outhwaite et al. (2022) called for learning apps to be evaluated in different settings, including the home environment, to enhance understanding of how they might address educational challenges faced in different contexts.

This chapter investigated how educational apps might address the global learning crisis by capitalising on the unique GLXP competition that directly compared five learning apps with learners of a similar age and ability who were all out of school. By exploring features of apps designed to support the acquisition of foundational literacy and numeracy skills, that were trialled within the context of the same RCT, with out-of-school children aged 7-11 years in a low-income country, this study adds to previous app evaluation frameworks that have focused on learning in one domain (e.g. maths) with early years or preschool children (Outhwaite et al., 2023; Kolak et al., 2021, respectively).

Current study

As there were clear winners and objective measures of learning gains from the GLXP, it is possible to directly compare embedded features of the finalist apps and draw inferences about which features are most effective at supporting learning for out-of-school children. To do this, a comparative judgement task was conducted, in which 41 naïve participants were asked to compare the five finalist apps on the 15 app features listed in Table 4 that have been proposed to support foundational learning. Comparative judgement was chosen over other methods to avoid potential inconsistencies that can occur in use of absolute judgement-based questioning (Kalton & Schuman, 1982), and biases that can occur (e.g., acquiescence bias) when using corresponding tools such as Likert Scales (e.g., Kim et al., 2021). Use of pairwise comparison allowed the five finalist apps and 15 pedagogical features to be clearly and consistently ranked both in a timely fashion and with no potential for ties (Marshall et al., 2020). Dichotomous ratings assessing if a feature was present within the apps (as in Outhwaite et al., 2023; Papadakis et al., 2018) were not used given their own potential to produce ceiling and floor effects (i.e., when a feature is present or absent in all apps). In contrast, the use of a comparative judgement task involved participants selecting an app that embodied a certain feature to a greater extent than another app. As such, relative judgements were able to order the extent to which different learning features were embodied within the five finalist apps.

This study mirrors previous work that has evaluated apps using a framework of features, but, for the first time, the learning gains achieved from the GLXP allow us to identify which features are most successful in facilitating children's learning of foundational skills in low-income community settings, by identifying key features of the two winning apps. To achieve this, the following questions were investigated:

1) Which features characterise each of the five finalist learning apps used in the GLXP?
2) Do the five finalist learning apps of the GLXP, that have been shown to support positive learning outcomes with out-of-school children within the same RCT, share features?

Method

Design

A within-participants experimental design was adopted in which all participants completed 165 trials (150 experimental trials and 15 repeated trials to assess consistency of response) of a two-alternative forced-choice comparative judgement task (a comparison method as outlined in Pollitt, 2012). Each trial required participants to judge which of two apps they were presented with were strongest on a particular feature (as listed in Table 4), such as *direct instruction* or availability of *autonomous learning*. In this manner, each of the five finalist apps was iteratively compared against each of the others. This gave rise to a "comparative judgement" score per app, which reflected the number of times the app was chosen over another in reference to a particular feature, with scores ranging between zero (never favoured) and four (always favoured) per app for each of the 15 features investigated. The order of comparative judgement trials was randomised across participants.

Prior to data collection a pilot was conducted with five participants, after which minor adjustments were made to refine this protocol based on researcher observations and participant feedback, such as presenting the full names of the apps throughout the task for clarity (e.g., CCI-School House instead of CCI). Analyses of the pilot data showed the task to be reliable and valid. All data was collected in a computer laboratory at the University of Nottingham. Ethical approval was granted by the School of Psychology Ethics Committee. Consent was obtained from all participants in line with the British Psychological Society guidelines.

Participants

Forty-one participants took part in the comparative judgement task described above. Participant ages ranged from 18 to 38 years (M= 24.14, SD = 4.51), with 27 females, 13 male, and one non-binary participants. All participants were residents of the United Kingdom and either currently or previously enrolled in Higher Education. Participants were recruited through the School of Psychology at the University of Nottingham using opportunity sampling via email invitation, posters, flyers, and word of mouth. Each participant was provided with an inconvenience allowance of £20 for taking part in the study.

All participants were blind to the study aims and had no a priori knowledge of either the results of the GLXP competition, or the five apps used, as confirmed by the researcher before each session. Rather than recruiting educational experts to this study, participants were non-experts thus mitigating risks associated with participants having different levels of background knowledge and expertise (Bramley, 2007).

Apparatus and Materials

Five Google Pixel C tablets with a 10.2” screen with a resolution of 2560 x 1800 pixels were used, one for each participant in any one session. Viewing distance was not controlled. The tablets were the same as those used during the GLXP competition with out-of-school children in Tanzania, and the five finalist apps were installed onto each tablet (see Chapter 2 for a brief description of each of the apps). The tablets enabled participants to access and interact with the five finalist apps before making comparative judgements. The same version of the apps used in this study were those used in the GLXP competition, where possible. Participants were made aware of any minor differences in app versions before completing the comparative judgement trials but were instructed not to base judgements on these differences. Specifically, the version of onebillion used in this study included a teacher loading page and the version of Robotutor used in this study did not block access to other apps, as in the XPRIZE version.

Participants used their own headphones when exploring the apps; these could not be provided due to COVID-19 precautions that needed to be followed. Pen and paper were provided to participants so that they could make notes. Each session started with a presentation delivered by the researcher that explained the GLXP, the aims of the current study and any information about deviations of the apps used in the study compared to those used in the GLXP, as described above. The session structure was outlined with instructions for each step. The results of the GLXP were not revealed at this time to avoid participant bias. All participants were provided with a booklet containing definitions for the 15 app features, framed as questions, as listed in Table 4. Some of the feature definitions were taken from specific papers or models (e.g., definitions from the Four Pillars of Learning framework; Hirsh-Pasek et al. 2015), while others were synthesised from multiple papers of app learning.

Qualtrics software, version February 2022, was used to present the comparative judgement trials and record participant responses. For each of the 15 features, there were ten possible app pairings, resulting in 150 distinct trials. Simple syntax was used to randomise question order and avoid bias. For example, when making judgements on the app feature of 'Gamification' participants were asked: '*Which app has more gamification? onebillion or Robotutor?'* Participants were required to click on the app they felt had the most of that feature.

To maintain attention and check for active participation throughout the task, after every 30 trials, participants were required to answer an unrelated two-alternative forced choice question, for example *'What season is it right now? Spring or Autumn?'*. Furthermore, for each of the 15 app features, one additional trial was repeated, but the app choice was reversed. For example, the question on Gamification given above was changed to *'Which app has more gamification? Robotutor or onebillion?* This was to determine consistency of participants' responses throughout the task.

Procedure

Participants contacted the researcher via email if they were interested in taking part. They were provided with an information sheet and were then invited to take part in a 2-hour session held at a time and date that was convenient for them. Each session had two to five participants, with most comprising five participants. At the beginning of each session, participants were welcomed, and the researcher presented the introductory information and task instructions, which took approximately 10-15 minutes. Participants were then asked to spend five minutes reading the feature definitions carefully and were given the opportunity to ask any questions for clarification.

Participants were shown how to use the tablets to access and interact with the five finalist apps. They were then asked to spend unstructured time exploring each app to familiarise themselves with the app features. During this process, it was recommended that participants make notes of anything positive, negative, or different that could support them in making the comparative judgements. They were instructed to spend a minimum of 25 minutes on this activity, 5 minutes per app, but assured that they could spend longer familiarising themselves with the apps if they wanted to. Participants were asked to engage with a range of literacy and numeracy tasks within each app and to give both correct and incorrect answers to experience the range of feedback provided. The researcher was on hand to answer questions and troubleshoot any technical issues.

When participants felt sufficiently familiarised with the apps, the researcher checked they understood the task, answering any remaining questions they might have. Participants then performed the comparative judgement task independently, with no group discussion. Participants were instructed to work at their own pace and reminded they could refer to the apps at any point when making their judgements, allowing further, targeted exploration if needed, to enable them to make informed decisions. Participants actively engaged with the apps whilst completing the comparative judgement task. Once all participants had completed the task, the researcher thanked them for taking part and answered any outstanding queries. Sessions lasted between 1 hour 40 minutes and 2 hours and 10 minutes, with an average duration of 1 hour and 55 minutes. Consent forms, debriefs, and withdrawal statements were provided during the session in accordance with ethical guidelines.

Data Analysis

Inclusion criteria for data analysis required participants to answer 100% of the unrelated attention questions and 80% of the ‘retest’ questions accurately. All participants achieved this and consequently no participants were excluded from the analysis.

Data was first analysed to assess the reliability of the comparative judgements: responses to the 15 repeated trials were compared to the corresponding experimental trial response for each participant. Internal consistency was determined by calculating the number of times participants gave the same response across repeated and original experimental trials, from which a group mean consistency was determined. Internal consistency for participant responses across the 15 repeated trials was high, with a group mean score of 13.83 (SD = 1.16; 92.2% accuracy), demonstrating comparative judgements were reliable. Fleiss’ kappa was then conducted to determine the agreement between participants for all experimental comparative judgements. Results revealed moderate agreement between participant judgements, $\kappa = .565$ (95% CI, .562 to .568), $p<.0001$.

Data was then analysed to determine if the comparative judgements were valid: responses to ‘Direct Instruction’ were used as a measure of face validity as the definition of this feature requires an app to have an avatar that guides and instructs the user (see Table 4), and hence can be objectively assessed. If the apps that include an avatar (i.e., onebillion, KitKit and CCI) were chosen more frequently as demonstrating ‘direct instruction’ than the apps that did not include an avatar (i.e., Chimple and RoboTutor), face validity will be confirmed. Accordingly, a binomial test of probability was conducted using the sum frequency count for 'direct instruction' for the three apps that have a clear avatar compared to the two apps that do not use an avatar. The sum frequency of responses to the 'direct instruction' feature for the three avatar apps was 333 compared to 77 for the two apps with no avatar. A binomial test revealed that the apps with an avatar were chosen significantly more than chance, $p<.001$, demonstrating the comparative judgements made were valid.

Data analyses were then conducted to address each of the research questions posed. For each of the 15 app features examined, participants made 10 comparative judgements, resulting in a maximum frequency count of four per app. As data was not normally distributed, non-parametric tests were conducted. Results are reported at a 2-tailed level of probability. Statistical analyses of the results of the experiment were conducted using Jamovi version 1.6 or Python version 3.10.4.

1) Which features characterise each of the five finalist learning apps used in the GLXP?

To identify features that characterised each of the different learning apps, a series of Binomial tests of probability was conducted using sum frequency counts for each feature per app, across the 150 experimental trials. Results indicated app features that were chosen significantly above or below chance.

2) Do the five finalist learning apps of the GLXP, that have been shown to support positive learning outcomes with out-of-school children within the same RCT, share features?

To determine if pedagogical features were common across the five finalist apps a series of ten Spearman's Rank correlation tests were conducted using sum frequencies per app across each of the 15 features examined through the 150 experimental trials. Bonferroni correction was applied to allow for multiple comparisons; adjusted significance level = 0.0125.

Results

1) Which features characterise each of the five finalist learning apps used in the GLXP?

Table 5 reports sum frequencies of the 15 pedagogical features for each of the five finalist learning apps. Green cells with bold text indicate results from the Binomial tests of probability where an app feature was favoured by participants at a level significantly more than chance; red cells with bold text indicate results for app features that were chosen significantly less than chance (at a 5% significance level). For each app, rank order of features is provided in parenthesis.

Table 5

Binomial test of probability results for each of the five finalist learning apps

App Feature	**Sum Frequency Count (total across apps = 410; chance = 82)** **Significance of Binomial Test**				
	onebillion	**Kitkit**	**CCI**	**Chimple**	**Robotutor**
Direct instruction	**144 (1)**	84 (13)	**105 (3)**	**57 (13.5)**	**20 (9)**
	<.001**	0.844	0.007**	0.002**	<.001**
Autonomous learning	**142 (2)**	**109 (6)**	**62 (13)**	82 (4)	**15 (13)**
	<.001**	0.002**	0.013*	1	<.001**
Curriculum links	**140 (3)**	68 (15)	98 (6)	**63 (11)**	**41 (1)**
	<.001**	0.966	0.059	0.019*	<.001**
Retrieval-based learning	**132 (4)**	98 (8)	92 (7)	**60 (12)**	**28 (7)**
	<.001**	0.059	0.242	0.006**	<.001**
Motor skills	**131 (5)**	**110 (5)**	**46 (15)**	90 (3)	**33 (4)**
	<.001**	0.001**	<.001**	0.353	<.001**

App Feature	**Sum Frequency Count (total across apps = 410; chance = 82)** **Significance of Binomial Test**				
	onebillion	**Kitkit**	**CCI**	**Chimple**	**Robotutor**
Task structure	**130 (6)** <.001**	**124 (2)** <.001**	85 (9) 0.749	**53 (15)** <.001**	**18 (11)** <.001**
Feedback	**130 (7)** <.001**	92 (9) 0.242	88 (8) 0.493	73 (8) 0.293	**27 (8)** <.001**
Social interaction	**130 (8)** <.001**	86 (12) 0.658	**108 (1)** <.001**	74 (7) 0.355	**12 (14)** <.001**
Active Learning	**129 (9)** <.001**	83 (14) 0.941	**100 (5)** 0.034*	69 (9) 0.118	**29 (6)** <.001**
Meaningful Learning	**127 (10)** <.001**	91 (10) 0.294	**102 (4)** 0.019*	**57 (13.5)** 0.002**	**33 (4)** <.001**
Engagement	**122 (11)** <.001**	**116 (4)** <.001**	79 (10) 0.766	77 (6) 0.584	**16 (12)** <.001**

App Feature	**Sum Frequency Count (total across apps = 410; chance = 82)** **Significance of Binomial Test**				
	onebillion	**Kitkit**	**CCI**	**Chimple**	**Robotutor**
Language demand	**117 (12)**	**107 (7)**	72 (11)	81 (5)	**33 (4)**
	<.001**	0.003**	0.239	0.961	<.001**
Personalisation	**100 (13)**	**133 (1)**	**107 (2)**	**65 (10)**	**5 (15)**
	0.034*	<.001**	0.003**	**0.037***	<.001**
Gamification	79 (14.5)	**121 (3)**	**59 (14)**	**132 (2)**	**19 (10)**
	0.766	<.001**	0.004**	<.001**	<.001**
Free play	79 (14.5)	88 (11)	68 (12)	**140 (1)**	**35 (2)**
	0.766	0.493	0.091	<.001**	<.001**

Note. **p<.01, *p<.05. The five finalist apps are ordered according to overall learning gains achieved across the 15-month GLXP field trial, with onebillion and KitKit being joint winners (XPRIZE, 2019). Binomial tests were conducted where: n = total number of judgements (trials) made per feature across the study, k = observed sum frequency count for the chosen app per feature, *p* = 0.2 [the probability that the chosen app will be selected on any particular trial], and q = 0.8 [the probability that the chosen app will not be selected on any particular trial].

As shown in Table 5, for the onebillion app (a joint winner of the GLXP), participants judged 13 out of the 15 app features examined as more present than for competitors at a level significantly greater than chance, with only Gamification and Free Play being selected at levels expected by randomness. In contrast, participants judged the features of the Robotutor app (which produced the lowest overall learning gains in the GLXP field trial) as being present significantly less than chance across all 15 features examined. This suggests that compared to other apps, the features of Robotutor were perceived as less present by participants.

2) Do the five finalist learning apps of the GLXP, that have been shown to support positive learning outcomes with out-of-school children within the same RCT, share features?

As shown in Table 6, Spearman's rho correlation coefficients were calculated to determine the relationship between apps across the 15 pedagogical features examined. Only one correlation was found to be significant when Bonferroni correction was applied for multiple comparisons at an increased significance level of .0125. A strong negative correlation was found between two of the runners up, Chimple and CCI, $r_s(39) = -0.688$, $p = .005$, demonstrating these apps differed significantly to one another in the presence of pedagogical features.

Table 6

Spearman's rho correlation matrix between apps across the 15 features examined

		OB	KK	CHI	CCI	RT
onebillion	Spearman's rho	—				
	p-value	—				
Kitkit	Spearman's rho	-0.366	—			
	p-value	0.180	—			
Chimple	Spearman's rho	-0.420	0.173	—		
	p-value	0.119	0.537	—		
CCI	Spearman's rho	0.165	-0.418	-0.688 **	—	
	p-value	0.557	0.123	0.005	—	
Robotutor	Spearman's rho	0.000	-0.520	0.070	-0.244	—
	p-value	1.000	0.047	0.804	0.381	—

Note. **p < .0125

Discussion

The aim of this chapter was to identify pedagogical features of interactive apps that are effective in supporting learning of foundational skills for out-of-school children in low-income settings. Here, for the first time, findings were reported from a comparative judgement task of the five finalist apps used in the GLXP that directly links app features to learning outcomes established over a 15-month field trial with out-of-school children in Tanzania. Two key findings were revealed.

First, across the five finalist apps, results showed that only the joint winners of the GLXP displayed pedagogical features that were significantly present, as judged by participants, more frequently than chance. In contrast, for the three runners-up, participants judged some pedagogical features (or all features in the case of Robotutor) as significantly less present than chance. Accordingly, these results provide evidence that the six pedagogical features shared by onebillion and KitKit – joint winners of the GLXP – were strongly present in these apps to support learning and included: *Autonomous Learning, Motor Skills, Task Structure, Engagement, Language Demand,* and *Personalisation*. Three of these features – *Autonomous Learning, Motor Skills*, and *Language Demands* – are centred on app accessibility which has been highlighted prominently within the literature (e.g., Gulliford et al., 2021; Lynch et al., 2021; Pitchford, 2023). Crompton et al. (2021c) highlighted accessibility to be a fundamental barrier to education in low- and middle-income countries (LMICs). The current results support prior research that emphasises the importance of pedagogical app features that enhance accessibility in supporting the acquisition of foundational skills.

Second, results showed no significant positive correlations between the five finalist apps deployed in the GLXP across the 15 features examined in this study, emphasising the variation in the presence of features across the applications. Only one correlation was significant, and that was a strong negative correlation, demonstrating a high degree of dissimilarity between the apps produced by Chimple and CCI – each targeting distinctly different sets of features to engage users. This variability is somewhat expected, due to the lack of direction given to app developers in the GLXP but is beneficial in the comparisons it allows us to make. The diversity in approaches adopted by the five teams regarding what they thought would constitute an effective learning app in this context is notable. For the joint winners of the GLXP, onebillion and KitKit, whose apps resulted in the highest learning gains over the 15-month field trial (XPRIZE, 2019), a weak negative correlation was found across app features, but again this was not significant, reflecting the dissimilarity in the presence of pedagogical features between the two apps. This variation between the two winning apps suggests that foundational skills can be supported by interactive apps that implement a range of pedagogical features. However, a core subset of features– specifically *autonomous learning, motor skills, task structure, engagement, language demand,* and *personalisation* – was strongly present in both apps.

It is not surprising that autonomous learning was strongly present in apps that supported foundational skills in this study, as the ability to learn autonomously was almost necessitated. The children in the GLXP were not provided with formal instruction beyond the app. This corroborates previous research proposing learner autonomy is critical to improving children's motivation, reflective engagement, and educational outcomes (Lan, 2018). Apps deployed in out-of-school settings should prioritise ease of navigation through the content to encourage and motivate independent learning by novice users. This is important even when installing tablets in school settings within LMICs, as there are often low teacher-to-pupil ratios and large class sizes (Jordan et al., 2021). Hence, apps that encompass pedagogical features that promote independent learning may be preferable for children to maximise their learning experience.

The results also highlighted the presence of features related to accessibility, such as motor skills and language demand, as important factors in fostering positive learning outcomes. Developers need to consider the level of motor skills required for children to interact effectively with their apps: too high a level of precision may have a negative impact on accessibility, especially for children with physical disabilities who may not possess the fine motor skills required to interact with the app content (Gulliford et al., 2021; Pitchford et al., 2018). Similarly, the language used in the app needs to be appropriate for the child's developmental age, which could be highly variable in LMIC out-of-school contexts. Language proficiency has been shown to correlate with children's learning with interactive apps (Gulliford et al., 2021; Outhwaite et al., 2020) so developers should keep the language as simple as possible to enhance the reach of their apps.

The other three app features strongly present in the joint winners of the GLXP were task structure, engagement, and personalisation. The importance of task structure has been highlighted in previous research with app use in primary school settings, due to the complementary relationship it has with the curriculum and use of reinforcement when learning new topics (Gulliford et al., 2021). Likewise, personalisation (with programmatic levelling), in combination with feedback, has been suggested to maximise learning outcomes when considering literacy app design (Vanbecelaere et al., 2020b) and maths app interventions (Outhwaite et al., 2023). For an app to have high educational quality, it should also support a child's engagement in the learning process, as engagement and learning "go hand in hand" (Raymer, 2013), using contingent interactions, extrinsic and intrinsic motivation (Hirsh-Pasek et al., 2015).

Interestingly, free play as a feature did not significantly characterise the two winning apps of the GLXP, despite previous research suggesting it plays an important role in exploration and autonomy (Hirsh-Pasek et al., 2015). Play-based learning (or guided play) has recently been shown to be successful in school settings (within a developed context), suggesting that play can be valuable if guided by a teacher with a learning objective, by balancing exploration and instruction (Skene et al., 2022). The importance of task structure highlighted in this study corroborates Skene et al. (2022) and demonstrates that in *out-of-school* settings, a certain level of structure is required within an app to direct a child through its content, and to promote positive learning outcomes where there is no formal teacher available to guide the learning process.

Limitations

A potential limitation of this study is that participants were based in the United Kingdom and had no experience or affiliation with Tanzania, where the GLXP was undertaken. Thus, participants were unfamiliar with the context in which the apps were deployed, which might be particularly pertinent when judging features such as ‘meaningful learning’ and ‘curriculum links’ for children in Tanzania. However, the apps investigated in this study are used in other countries and contexts, such as the UK and Brazil, in which demonstrable learning gains have been achieved (Outhwaite et al., 2020), suggesting that the presence of these features is not country specific.

Conclusion

This chapter has identified six pedagogical app features – *autonomous learning, motor skills*, *task structure*, *engagement, language demand,* and *personalisation* – that were judged to be significantly present in apps associated with the effective learning of foundational skills in low-income community settings. This combination of app features appears to be key to ensuring the optimal effectiveness of learning apps deployed in LMICs, where spending budgets for education are extremely limited. Chapter 4 (Domain and Gender App study) investigates this further by exploring the relative importance of the 15 app features and examining differences in the impact of individual features based on the domain (i.e. literacy/numeracy) or gender that they are targeting. Future studies should assess the reliability of this combination of pedagogical features in other educational apps and settings to evaluate their adoption in different contexts.

The results of this chapter should inform the pedagogical design of educational apps, particularly for use by children of primary school age in LMICs and should be useful to governments, educators, and parents when deciding on educational apps to support the acquisition of foundational skills, especially with out-of-school children. This is crucial considering that 244 million children worldwide are estimated to be out of school (UNESCO, 2022a), and 10% of global spending is purported to be wasted on ‘poor education’ that perpetuates significant inequalities in access to and provision of quality education that is failing to produce the desired learning outcomes for children (UNESCO, 2019). With the combination of pedagogical features identified in this chapter, well-designed educational apps could start resolving this global foundational learning crisis.

Chapter 4: Relative Importance of App Features across Domains and Gender

Background

Building on the results of Chapter 3, this chapter explores the individual impact of the 15 different app features on foundational literacy and numeracy outcomes for out-of-school girls and boys in Tanzania. As highlighted in Chapter 1, while surprisingly, there is no significant gender gap in Tanzanian enrolment rates, there is still an attainment gap in learning outcomes that favours boys (Groeneveld & Taddese, 2020; Al-Samarrai & Tamagnan, 2019). Therefore, this would suggest that the education provision is not as impactful for girls, and further research is needed to determine what may facilitate their learning in low-income, low-resource settings. As research has shown that learning apps could be helpful in facilitating learning for girls (e.g. Pitchford et al., 2019), this exploration of app features and their relationship with gender and domain is crucial to understanding and enhancing technology-driven education in areas where access to conventional education is scarce, and gender disparities may impact foundational learning. This study utilises a completely novel research methodology, combining machine learning methods and inferential statistics, to explore the differential impact of app features on learning outcomes, dependent on domain and gender, in a low-income, out-of-school context for the first time.

App features and learning outcomes

Attempts to link app features to learning outcomes are scarce. As explored in Chapter 3, Outhwaite et al. (2022) provided a critical step towards capturing this relationship in maths with a systematic review and content analysis of maths apps. The systematic review had a global focus, including high- and low-income research. A qualitative comparative analysis of apps found on the UK Apple and Google Play stores was then utilised to identify specific app design features, such as scaffolding and personalisation, that maximised children's learning outcomes in maths. This research underscored the importance of linking app features directly to learning outcomes, an objective that continued in Chapter 3 but with Tanzania as the direct focus, providing a different lens to the Westernised, high-income focus of the UK.

In Chapter 3, Outhwaite's work was built upon by comparing the app features of the five finalist apps from the GLXP, employing a comparative judgement task to assess 15 pedagogical features. Six pedagogical app features were identified as being prominent in both apps that produced the greatest learning outcomes in the XPRIZE competition (i.e. the two winning apps, onebillion and KitKit School): autonomous learning, motor skills, task structure, engagement, language demand and personalisation. While these insights are valuable, they do not distinguish between the impacts of different app features on literacy outcomes independent of numeracy outcomes and, therefore, may not fully capture the complex relationship among app features, complex inter-feature interactions and their contributions to different domains of foundational learning.

Investigating app features concerning independent literacy and numeracy outcomes has not yet been studied at the level of individual learners. This could be crucial in identifying app features that are most successful in predicting whether individual out-of-school learners will achieve foundational literacy and numeracy skills.

Gender differences

Furthermore, understanding evidenced gender disparities in educational outcomes is essential, particularly in low-income countries, where distinct differences in how boys and girls access education may influence their engagement with tools like educational apps. As highlighted in Chapter 1, it is estimated that there are 6.7 million out-of-school children in Tanzania, where girls are among those most at risk (UNESCO, 2016). Although initial school enrolment rates are similar across both genders, retention falls substantially for girls due to a multitude of reasons such as poverty, marriage, pregnancy and having traditional gender roles to fulfil within their households (UK Aid, 2021; Awinia, 2019). As a result, there are also significant gender disparities in learning outcomes at primary school, which widen further after the transition to secondary school, with girls generally receiving lower examination results than boys (Awinia, 2019; Al-Samarrai & Tamagnan, 2019). This exacerbates the dropout rates further, with a 1:2 girl-to-boy ratio by the time students reach upper secondary (Pezzulo et al., 2022). Accordingly, governments worldwide, including the UK, have focused on improving girls' education to equalise opportunities and reduce barriers to girls achieving their full learning potential.

Can girls learn with apps in low-income countries?

Despite girls from low-income countries typically facing more difficulties accessing education than boys, research has shown that girls can enjoy and benefit from education just as much as boys when provided with equal access, both in a traditional school setting and with a learning app in an out-of-school setting (Miheretu, 2019; XPRIZE, 2019). Unlike traditional schooling, apps offer flexibility as they can be used at any time, meaning their use could be accommodated alongside the expected household responsibilities of girls (UK Aid, 2021). Therefore, educational apps may provide a valuable tool to enhance girls' education and reduce the attainment gap, especially if the app features are known to promote and not inhibit engagement and app learning.

Gender differences in app features

Within conventional school settings in low-income countries, there have been notable gender differences when learning foundational skills, with boys typically outperforming girls in numeracy but girls having a higher reading level than boys (UNESCO, 2015; OECD, 2016). However, it is essential to note that these differences are domain-specific rather than specific to app features. How girls and boys engage with these features might vary depending on their strengths and engagement in each domain. This variation could offer insights into tailoring apps to better suit gendered learning preferences, enhancing the effectiveness of apps in addressing any disparities in literacy and numeracy skills.

Very little previous research has explored gender differences across app features; one exception is gamification. Research into gamification has explored gender differences across subject domains, ages, and game styles. Findings suggest that girls are less interested in the competitive elements of gamified learning than boys are, with boys more likely to compete with others, while girls tend to focus on completing their assignments during gameplay (Kickmeier-Rust et al., 2014; Admiraal et al., 2014). Boys were also more interested in rewards and achievements within the app, although there were still no significant differences in the level of achievement and amount of learning rewards received (Yang & Quadir, 2018). While a helpful insight into potential gender differences, this research was based on self-reported enjoyment and motivation (e.g. Kickmeier-Rust et al., 2014) or just comparing differences in learning outcomes between genders after using one gamified app (e.g. Yang & Quadir, 2018). To understand if there are gender differences in the impact of different app features, direct comparisons of different apps should be made in relation to known learning outcomes, allowing an exploration of gender differences between apps that embody different amounts of each app feature (e.g. some apps are more focused on gamification than others, as demonstrated in Chapter 3).

Research focusing on gender differences for the other key app features remains sparse. However, within the gamification literature, it was found that boys are more interested in interactivity than girls, suggesting that an active learning approach within an app may appeal to boys (Yang & Quadir, 2018). It was also found that while boys appreciate competitive gameplay, girls tend to place more importance on the discovery and exploration enabled by free play (Admiraal et al., 2014) and the feedback that they receive within the game (Kickmeier-Rust et al., 2014).

While there are insights into gender differences for gamification, similar exploration is lacking for the other fourteen learning apps studied in Chapter 3. Additionally, while the gamification research is a helpful start, it is focused on different settings (high-income schools rather than low-income out-of-school) and demographics (e.g. secondary students). The gender differences that have been found focus on preferences and opinions rather than tangible learning outcomes and the research was not based in Tanzania. Given these limitations, it would be highly beneficial to explore whether there are clear gender differences in the impact of key app features on foundational learning outcomes for out-of-school children in Tanzania so that app developers can tailor the technology to mitigate gender differences.

Current study

This chapter aimed to address the aforementioned significant gaps in the research. Previous studies have provided valuable insights into app features and their general impact on learning. However, there needs to be more research at the granular level of individual learners, particularly in low-income contexts. This could be crucial in identifying app features that are most successful in predicting whether individual out-of-school learners will achieve foundational literacy and numeracy skills. Therefore, a primary objective of the current exploration was to determine the most influential app features in predicting whether children will improve their foundational literacy and numeracy skills among out-of-school children in Tanzania.

Furthermore, the app features and their differing impact on girls and boys remain underexplored, so another aim was to provide a more comprehensive understanding of how gender might influence the effectiveness of the 15 educational app features in out-of-school contexts. In combination, these two objectives will look across and within gender and domain to identify the vital features for successful educational technology, building on findings from Chapter 3.

To address these pivotal questions, new data analytic methods are needed to allow analysis of the data at an individual level and untangle the complex relationship between app features and learning outcomes. Traditional data analysis methods face challenges that impact the interpretation and generalisability of the resulting models, such as multicollinearity, assumptions of linearity and overfitting (Lavelle-Hill et al., 2021). Correlations between variables lead to potentially important variables being excluded, altering explanations for the outcome variable being investigated. Further, traditional regression using small datasets leads to a considerable risk of overfitting, which undermines the resulting model interpretation.

In contrast, machine learning techniques utilise alternative methods to address these problems, such as cross-validation, model parameterisation and out-of-sample testing, where the model's performance is assessed using predictions on a held-out test set. This approach is, therefore, particularly suited to evaluating the relationship between the 15 independent app features and learning outcomes, especially when focusing on gender and domain differences. Out-of-sample prediction can also be preferable for research with real-world applications such as education, as this method can be used to predict future data points, and past experiences have been shown to predict future behaviour (Lavelle-Hill, 2020). Big data for educational research is still in its infancy, so conducting this research using machine learning techniques may also provide insights about its suitability in this context and how these methods may be utilised further in future educational applications (Yu, 2020).

This research employed a novel mixed-methodological approach, combining machine learning with traditional inferential statistics. Machine learning methods were used first to construct prediction models that allowed us to see which app features were most predictive of learning improvement across genders and domains and what direction their influence was in (i.e. did the app feature positively influence the learning outcomes?) ANOVAs and post-hoc t-tests then enabled rigorous testing to determine which differences in the impact of app features across genders and domains were statistically significant. The combination of methodologies allows for identifying the most influential app features and understanding the complex interactions within and across genders and domains for all app features. These research methods allowed us to address the following primary research questions:

1. Which app features most predict improvements in literacy and numeracy outcomes for out-of-school girls and boys in Tanzania following an Edtech intervention?
2. Are there significant differences in the impact of individual app features on literacy and numeracy outcomes for out-of-school girls and boys?

The findings from this study could potentially influence the development and implementation of EdTech in low-income contexts by identifying the app features that effectively support literacy and numeracy development while taking any gender differences into account. This could inform the creation of more considered, impactful tools for remote education and help governments and implementing agencies make informed decisions about which apps to use in out-of-school contexts, ultimately informing educational policy to help reduce existing educational inequalities.

Method

This investigation explored the comparative importance of 15 app features examined in Chapter 3 concerning learning improvements in literacy (gains in EGRA scores over time) and numeracy (gains in EGMA scores over time) for boys and girls following the EdTech intervention. The primary focus was identifying which app features are most important in predicting whether children will improve.

Dataset

Two datasets taken from earlier chapters were used for this analysis. The first dataset consists of key demographic data from the 2041 children who took part in the GLXP competition, namely their age, gender, the app that they received and their raw improvement scores for both EGRA and EGMA assessments following the competition period (explained fully in Chapter 2).

The second dataset is the comparative judgement frequency scores taken from Chapter 3, which will be used to conduct the Bradley-Terry model and generate lambda scores for each of the 15 app features.

Once the lambda scores were generated (explained further in the Bradley-Terry model section), each of the scores for the 15 app features was then used to create a scoring column for each feature, joined to the child-level data in the first dataset based on which app they received, as the lambda scores were at the app level. All children allocated to the control group had missing values, as they did not receive any of the five apps used in the competition, so their data was removed because they had no experience with the app features. Therefore, the final dataset included 1680 data points, relating to all of the children that received one of the app interventions.

Generating lambda scores - the Bradley-Terry model

The Bradley-Terry model is a probability model commonly used to evaluate paired comparison data (Bradley & Terry, 1952). The model highlights which app features are preferred by comparing pairs of features, allowing them to be ranked based on user comparisons. This made it appropriate to further evaluate the comparative judgement frequency data introduced in Chapter 3, in which binomial tests of probability were utilised to generate a ranked order for each app's performance across the 15 app features, determining which apps scored significantly above chance for how many times their embodiment of the app feature was favoured by participants across the forced-choice task. onebillion came out as the strongest overall app across features, scoring significantly above chance for 13 out of 15 features.

In the current chapter, the outcome of these forced-choice comparisons- (e.g. Which app was better at providing direct instruction? onebillion or Robotutor)- was then used to calculate lambda (λ) scores for each item, representing the relative strength of each app feature found in each of the five apps. Lambda scores are calculated using the assumption that the probability $P(i > j)$ that app i is preferred over app j for a given feature is $\lambda i \lambda i + \lambda j$, where λi and λj are the lambda scores of apps i and j, respectively. Lambda scores were estimated using maximum likelihood estimation, ensuring the most accurate representation of the comparative judgement outcomes. The model assumed independent comparisons, and there were no ties due to the forced choice procedure employed in the initial comparative judgment task. In interpreting the lambda scores, high values indicate that an app feature is strong within the app, with low values (close to zero), showing that the app does not embody much of the app feature.

Data cleaning for modelling

Before analysing the dataset further, preparing the data for a classification analysis was essential. The primary goal was to determine whether girls and boys achieved or did not achieve significant improvement using the app. This was assessed using two dependent variables, the EGRA and EGMA improvement scores. The 'female' variable was coded as '1' and '0' for boys to facilitate the independent investigation of both genders. Within the final dataset, there are 631 girls and 1049 boys. Four classification tasks were conducted, one for each gender per dependent variable - these were Model 1 (girls' literacy), Model 2 (boys' literacy), Model 3 (girls' numeracy) and Model 4 (boys' numeracy).

The exploratory analyses of the raw improvement scores (across both genders), as depicted in Figures 6 and 7 for EGRA and EGMA, revealed notable deviations in the data distributions. Specifically, the EGRA dataset showed a skewness of 1.55 and a kurtosis of 1.41, while the EGMA dataset showed a skewness of 1.06 and a kurtosis of 0.77. These values indicate a highly positively skewed distribution for both EGRA and EGMA, with most children showing minimal improvement. Consequently, binary classification tasks were considered more suitable than regression for this analysis, with a binary outcome variable categorising children's performance with the app as 'high achievement' or 'low-achievement'.

Implementing a binary cut-off point to distinguish between 'high-achievers' and 'low-achievers' can reduce noise in the dataset, enhancing prediction accuracy (Ahmad, 2019). These cut-off points were determined by visually examining the histograms in Figures 5 and 6. Starting with higher bin counts and then gradually reducing them, this process allowed for a clearer identification of a sensible cut-off point, reflecting the point where a significant drop in frequency was observed, effectively separating the two groups for each dependent variable.

Figure 5

A histogram showing the raw EGRA improvement scores after the GLXP

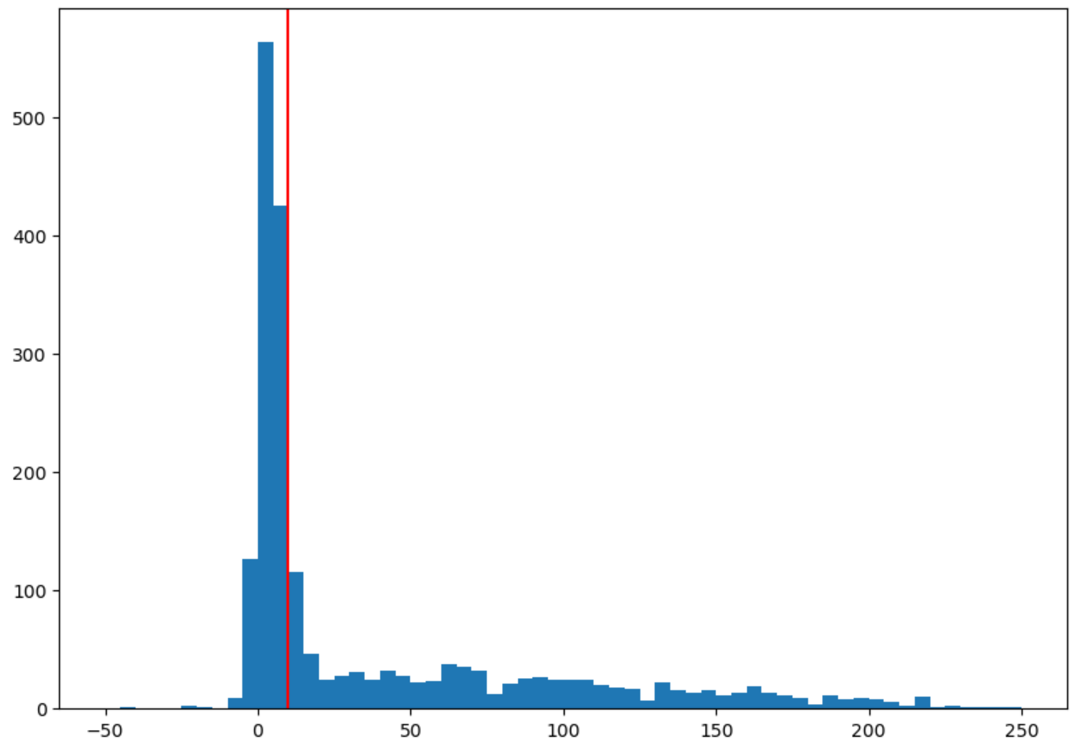
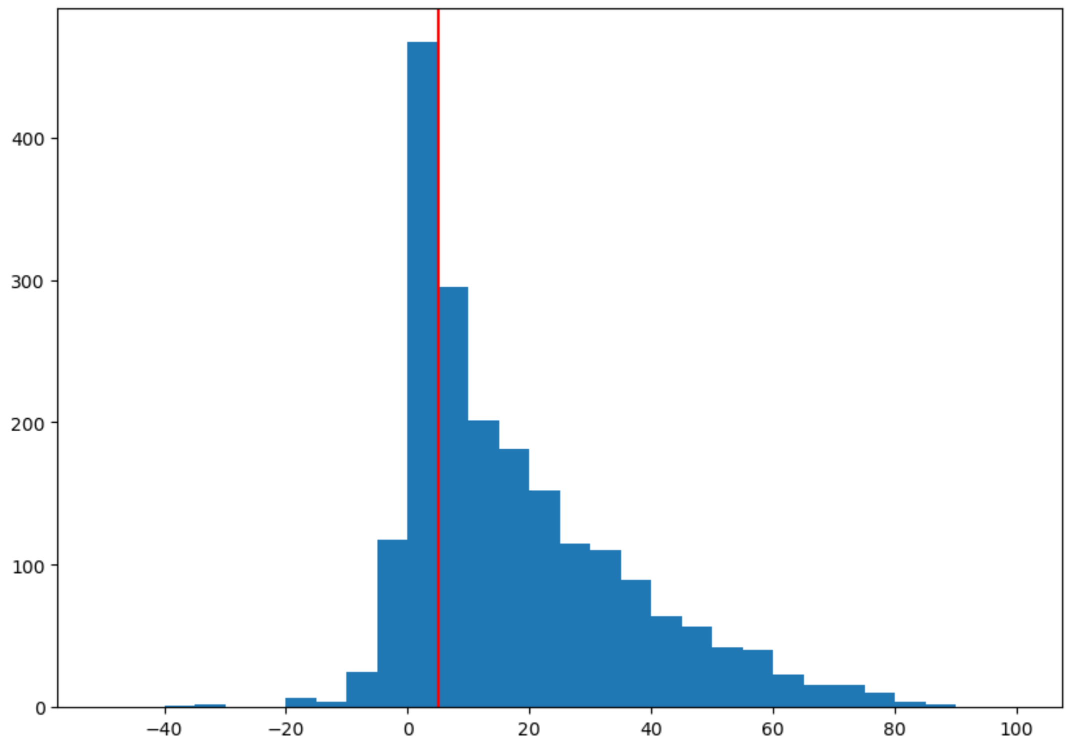


Figure 6

A histogram showing the raw EGMA improvement scores after the GLXP



The chosen cut-off point for EGRA (indicated by the red line on Figure 5) was a raw improvement score of 10 or below as low-achievers (coded '0'), and any score above that would be high-achievers (coded '1'), who surpassed the improvement threshold. This led to a reasonably balanced dataset for both girls and boys, with 309 low achievers (49%) and 322 high-achievers (51%) for girls and 551 low-achievers (52.5%) and 498 high-achievers (47.5%) for boys.

The cut-off point for EGMA (see the red line in Figure 6) was five and below for low achievers (coded '0'), and everything above was classified as high achievers (coded '1'). This led to two imbalanced datasets, a dataset with 186 low-achievers (29.5%) and 445 high-achievers (70.5%) for girls and 281 low-achievers (26.8%) and 768 high-achievers (73.2%) for boys.

Modelling

Four classification tasks were formulated, one for each gender per domain, to investigate the importance of app features that may predict whether girls and boys will improve their foundational literacy and numeracy skills following an EdTech intervention.

The chosen machine learning model was the Random Forest classifier**,** an ensemble tree-based method that learns and combines a series of individual decision trees to create the 'forest', significantly improving the model's overall accuracy. Random Forest is known for its high efficiency with large databases and is more robust to errors and outliers than other boosting methods, as well as slight imbalances in the data (Han et al., 2022; More & Rana, 2017). Additionally, Random Forest can handle non-linear data and interaction effects, making it suitable for capturing complex relationships found in educational data. Random Forest is also less prone to overfitting than alternative machine learning models, as the averaging process across multiple trees improves the model (Segal, 2004). Random Forest can provide valuable insights into feature importance by quantifying the impact of each feature on prediction accuracy, offering a clear understanding of which app features are most influential in predicting literacy and numeracy learning outcomes. This was particularly important for the aim of this investigation, which was to identify and rank the effectiveness of the different app features.

A dummy classifier can also be implemented to provide baseline accuracy scores for comparison against the Random Forest classifier. Dummy classifiers make basic predictions that ignore the input features and make predictors based on the observed dependent variable values (the ‘y’ values observed in the ‘fit’ argument; Scikit Learn, 2023). For a binary classifier, the dummy performance can be assumed to be 0.5, so this assumption was used to compare against the Random Forest performance. In terms of outcomes, a dummy performance of 0.5 means that the classifier would correctly identify the class of a child (low-achiever or high-achiever) 50% of the time, equivalent to a coin flip.

The core classification steps were all conducted four times - for Model 1 (girls' literacy), Model 2 (boys' literacy), Model 3 (girls' numeracy) and Model 4 (boys’ numeracy). Firstly, the data was split into a training and test set (80/20 split), with the training data used to train the model, allowing the model to make predictions based on the input data. The test set is then used to evaluate how well the model can perform on the unseen data, ensuring that the model can generalise to make new predictions and prevent overfitting the dataset (Brownlee, 2020). The model performance was assessed using a K-Fold cross-validation approach, splitting the sample repeatedly into training and validation sets, where K = 5, as is now common in machine learning analysis to balance the bias and variance in the test error rate (Nica-Avram et al., 2021). Grid searches were performed using the 80% training subsample to identify optimal hyperparameters.

The predictive performance of the Random Forest model was then measured on the held-out test set using the optimised parameters, evaluated by accuracy and F1 scores. The accuracy metric measures the ratio of correct predictions over total instances. Accuracy is often used as an evaluation metric in binary classification problems (Hossin & Sulaiman, 2015).

$$\text{Accuracy} = \frac{true\ positive + true\ negative}{true\ positive + false\ positive + true\ negative + false\ negative}$$

The combined F1 metric represents the harmonic mean between recall and precision values and is a good discriminator that can outperform accuracy for binary classification problems (Hossin & Sulaiman, 2015).

$$\text{F1} = \frac{2 * precision * recall}{precision + recall} \text{ where:}$$

$$\text{precision} = \frac{true\ positive}{true\ positive + false\ positive} \quad \text{and} \quad \text{recall} = \frac{true\ positive}{true\ positive + true\ negative}$$

Testing each model on this unseen data prevents the model from overfitting to the sample and ensures that the model generalises (Lavelle-Hill, 2020). This process was repeated 30 times, giving 30 scores for the final model performance (as seen in Engelmann et al., 2018). Repeating the training process multiple times and averaging the iterations provides a more reliable estimate of the model's performance, as it ensures that the results are robust to randomness; when using a single model run, it may be particularly high or low performing, making it difficult to generalise a singular result.

While Random Forests can effectively handle class imbalances in data reasonably well (More & Rana, 2017), the EGMA dataset was highly imbalanced towards the majority class (1), which in this case was the high-achievers, as there was a much higher percentage of high-achievers than low achievers. This imbalance could potentially bias the Random Forest and lead to decreased performance in distinguishing the minority class (Sun & Shen, 2022). The minority class (0; the low achievers in this case) was oversampled to match the size of the majority class to address this imbalance. This was randomised for each iteration of the model loop to ensure that different data was duplicated from the minority class each time.

The Random Forest model was then fitted to the complete dataset with the optimal parameters to identify the relative importance of each app feature. The evaluation metrics were checked to ensure no significant decrease in performance, and then analyses of variable importance were undertaken using SHAP (SHapley Additive exPlanations; Lundberg & Lee, 2017). SHAP values provide a measure of the average contribution of each feature to every prediction. They are frequently used to evaluate feature importance, making complex model outputs, particularly tree-based models, easier to interpret (Hong & Frias-Martinez, 2020). SHAP was chosen over permutation importance as importance is based on a magnitude of feature attributions, thus enabling an understanding of how different feature values affect the predicted outcome and, ultimately, which features are more important than others. SHAP values represent a measure of the average contribution of each app feature to every model prediction, allowing evaluation of feature importance.

The entire classification model and SHAP pipeline was looped 30 times for each of the four models, providing 30 scores for the final model performance and 30 SHAP values for each app feature. Conducting 30 SHAP analyses for each model significantly enhances the robustness of findings on feature importance, as each iteration of a machine learning model may yield a different order of feature importance due to the model variability. Averaging the SHAP values across the 30 runs allows an assessment of feature importance across 30 Random Forest models and the identification of consistent patterns in how features influence the outcome prediction, providing a more reliable and comprehensive assessment of feature importance. This allowed a ranked list of app features for each gender per domain to be generated based on the average SHAP values, allowing the most influential app features in predicting literacy and numeracy learning for boys and girls, respectively, to be determined.

Inferential analysis

Following the modelling, inferential statistics were used to investigate further the relationships between app features, domains (literacy and numeracy), and gender. This involved conducting 2 (Domain: literacy, numeracy) x 2 (Gender: girls, boys) ANOVAs for each of the 15 app features, using the raw SHAP values taken from each of the 30 classification loops to examine how app features interact with domain and gender. The key area of interest for this study was the interaction effects, as the aim was to identify if there was a significant relationship between the impact of domain and gender on learning outcomes.

Post-hoc t-tests were then used to explore the app features with significant interaction effects. Four post-hoc comparisons were made for each app feature with a significant interaction: girls vs boys in literacy, girls vs boys in numeracy, girls in literacy vs numeracy and boys in literacy vs numeracy. The first two comparisons are a between-group design, so one-sample t-tests were used, while the other two were a within-group design and used paired samples t-tests. This analysis enabled the understanding of how the features' impact varied across different groups and interactions.

ANOVAs and t-tests were chosen due to their ability to determine differences in means across groups and their significance, which is critical to understanding the impact of app features. ANOVAs were particularly suitable to analyse the factorial design (gender x domain), with t-tests ideal for making nuanced comparisons between and within gender and domain.

One of the key assumptions for ANOVAs and t-tests is that the data is normally distributed. The normality of the residuals was checked for each ANOVA using the Shapiro-

Wilks test, and any results where $p<0.05$ were considered to have departed significantly from normality. Despite this assumption, ANOVA is considered a robust test against violations of normality (Blanca et al., 2017), so the parametric 2x2 ANOVAs were reported regardless. For the app features that violated the normality assumption, the corresponding post-hoc tests were checked by conducting the equivalent non-parametric tests - Mann-Whitney for one sample t-tests and Wilcoxon Signed Ranks for the paired comparisons t-tests. If the parametric results aligned with the non-parametric results, then the parametric results were reported. When the results differed, the non-parametric tests were reported.

A key consideration for this study was to decide on an appropriate significance level. Given the multiple comparisons in this analysis, balancing the risk of Type 1 errors (false positives) and Type 2 errors (false negatives) was imperative. To address this balance, the significance level was set at 0.01. This was informed by the need to mitigate the increased risk of Type 1 errors associated with multiple comparisons while avoiding an overly conservative threshold that could lead to Type 2 errors, where true effects might be overlooked. The 0.01 was chosen as a compromise, being more stringent than the standard 0.05 but less conservative than the Bonferroni-adjusted level of 0.0033 (0.05/15). A more conservative adjustment, such as 0.0033, could potentially obscure significant findings, so 0.01 aims to reduce the likelihood of false positives while still allowing for the detection of meaningful results. This 0.01 threshold is also recognised in medical research as a stringent yet practical alternative to the 0.05 significance level (Jafari & Ansari-Pour, 2019).

To complement the inferential tests, effect sizes were calculated for each result. Partial eta squared values ($\eta 2$) were used for the 2x2 ANOVAs, which measure the proportion of variance in the dependent variable - achievement in literacy or numeracy - that can be attributed to the independent variables (domain and gender; Lakens, 2013). For all parametric t-tests, Cohen's d was calculated, which estimates the standardised difference between the two group means (Cohen, 1988). The effect size was calculated using the rank-biserial correlation coefficient for the non-parametric Mann-Whitney and Wilcoxon Signed rank tests. This provides clear insights into the strength and direction of the relationship between the two groups.

All statistical analyses were performed using Python version 3.12.0 and SPSS version 29.

Results

Results from each stage of the analytical process are presented to address the primary research questions, looking at which features are most predictive of improvement in literacy and numeracy outcomes for girls and boys and whether there are significant differences in the impact of app features across genders and domains.

Bradley-Terry analysis

Table 7 displays the lambda scores calculated for each of the 15 app features per app, indicating their relative preference based on the Bradley-Terry analysis. This allows you to see which apps are stronger at embodying each chosen feature, with high scores meaning that the app was considered strong for that feature.

The results are consistent with the binomial findings in Chapter 3, in which the app produced by onebillion outperforms the other apps across many of the 15 app features. The lambda scores for 12 of the 15 app features investigated were highest for the onebillion app compared to the other four apps that were tested in the GLXP, demonstrating that onebillion is the strongest app when considering the embodiment of the chosen features of importance. Of particular note is that the lambda score for autonomous learning was by far the highest for onebillion compared to the other learning apps, which was the pedagogical feature the GLXP sought to promote with out-of-school children. onebillion also scored highest for direct instruction and curriculum links, which in combination with autonomous learning, are the three highest lambda scores across the entire Bradley Terry analysis and were the top three ranked features for onebillion in Chapter 3. Kitkit School, the other winner of the GLXP, had a substantially higher score for personalisation than the other four apps, suggesting personalisation is a unique strength of this app.

Table 7

Lambda scores for each app feature

App Feature	App Lambda Scores				
	onebillion	**Kitkit School**	**CCI**	**Chimple**	**Robotutor**
Direct instruction	**0.593**	0.116	0.200	0.066	0.024
Autonomous Learning	**0.563**	0.219	0.072	0.126	0.020
Curriculum links	**0.551**	0.095	0.200	0.095	0.060
Retrieval-based learning	**0.475**	0.201	0.184	0.095	0.045
Motor skills	**0.446**	0.261	0.064	0.183	0.046
Task structure	**0.416**	0.352	0.144	0.064	0.024

App Feature	App Lambda Scores				
	onebillion	**Kitkit School**	**CCI**	**Chimple**	**Robotutor**
Feedback	**0.442**	0.187	0.191	0.129	0.050
Social interaction	**0.455**	0.148	0.263	0.113	0.020
Active learning	**0.438**	0.150	0.245	0.120	0.047
Meaningful learning	**0.426**	0.179	0.255	0.090	0.051
Engagement	**0.367**	0.316	0.150	0.140	0.027
Language demand	**0.346**	0.277	0.142	0.173	0.062
Personalisation	0.193	**0.460**	0.248	0.089	0.010
Gamification	0.118	0.327	0.071	**0.459**	0.024
Free play	0.126	0.151	0.101	**0.576**	0.045

Note. The highest lambda score for each app feature is highlighted in bold.

The lambda scores primarily inform the classification models, as they reflect the strength of each app feature across the five different educational apps and are important for understanding the relative feature influence on learning outcomes for literacy and numeracy.

Classification models

Four classification models were employed to evaluate how each of the 15 chosen app features influenced literacy and numeracy achievement for boys and girls following an Edtech intervention. This was followed by model interpretation using feature importance to determine the relative importance of each app feature.

Model performance

Table 8 reports the results of these investigations, showing the mean model performance for both genders for literacy and numeracy over 30 iterations on the held-out test set. Overall, the model predicted girls' performance with marginally greater accuracy for both domains than boys, suggesting that the chosen app features may be more influential for girls' learning improvement than boys.

Table 8

Summary of the mean Random Forest performance for literacy and numeracy improvement per gender

Performance	Literacy (EGRA improvement)		Numeracy (EGMA improvement)	
	Girls	**Boys**	**Girls**	**Boys**
Accuracy	**0.562**	0.553	**0.607**	0.579
F1 Score	**0.558**	0.534	**0.602**	0.579

Note. The highest score for each dependent variable is highlighted in bold for each measure (accuracy, F1).

As shown in Table 8, accuracy scores (ratio of correct predictions over total instances) and F1 scores (harmonic mean between recall and precision values) were extremely low for a binary classification task, with model performance being only slightly better than chance (which would be approximately 0.5). However, it is important to note that the models still achieved a performance level above that of a dummy classifier, which would predict outcomes at a 50% success rate by chance alone. This indicates that while the models' predictive powers are limited, they do offer some predictive advantage over a purely random classification, particularly in differentiating 'low' from 'high' achievers following an EdTech intervention.

Feature importance

Table 9 reports the averaged SHAP values for the different app features across the 30 classification models, showing each feature's impact on influencing whether a girl or boy will achieve literacy and numeracy improvement following the EdTech intervention. Figure 7 displays the distribution and variability of the 30 SHAP values for each app feature using four box plots, one for each gender per domain. This provides a visual representation of the dispersion of the SHAP values to complement the averaged values presented in Table 9.

Table 9

Mean (SD) SHAP values from the 30 Random Forest classification models and their overall rank in influencing the prediction of whether a child achieves foundational skills following an EdTech intervention

App Feature	**Literacy** Mean (SD) Rank order of magnitude		**Numeracy** Mean (SD) Rank order of magnitude	
	Girls	**Boys**	**Girls**	**Boys**
Retrieval-based learning	0.009 (0.003) 1st	0.007 (0.003) 4th	0.015 (0.005) 3rd	0.011 (0.003) 3rd
Engagement	0.009 (0.004) 2nd	0.008 (0.002) 1st	0.018 (0.005) 1st	0.012 (0.003) 2nd
Task structure	0.009 (0.003) 3rd	0.007 (0.002) 6th	0.016 (0.006) 2nd	0.012 (0.004) 1st
Language demand	0.007 (0.004) 4th	0.007 (0.003) 5th	0.014 (0.006) 4th	0.009 (0.005) 4th
Motor skills	0.007 (0.004) 5th	0.008 (0.003) 2nd	0.012 (0.005) 6th	0.012 (0.004) 5th

App Feature	Literacy Mean (SD) Rank order of magnitude		Numeracy Mean (SD) Rank order of magnitude	
	Girls	**Boys**	**Girls**	**Boys**
Personalisation	0.007 (0.004) 6th	0.001 (0.001) 15th	0.006 (0.003) 12th	0.006 (0.003) 10th
Autonomous Learning	0.006 (0.003) 7th	0.008 (0.003) 3rd	0.012 (0.004) 5th	0.008 (0.004) 6th
Social interaction	0.005 (0.003) 8th	0.003 (0.002) 8th	0.007 (0.004) 8th	0.006 (0.003) 8th
Feedback	0.004 (0.003) 9th	0.003 (0.001) 9th	0.007 (0.004) 9th	0.006 (0.002) 7th
Active learning	0.004 (0.002) 10th	0.003 (0.001) 12th	0.007 (0.004) 7th	0.007 (0.003) 11th
Meaningful learning	0.004 (0.002) 11th	0.003 (0.001) 10th	0.006 (0.003) 11th	0.006 (0.003) 9th

App Feature	Literacy Mean (SD) Rank order of magnitude		Numeracy Mean (SD) Rank order of magnitude	
	Girls	**Boys**	**Girls**	**Boys**
Direct instruction	0.004 (0.002) 12th	0.003 (0.002) 11th	0.007 (0.004) 10th	0.005 (0.003) 12th
Gamification	0.002 (0.001) 13th	0.002 (0.001) 14th	0.003 (0.001) 14th	0.004 (0.003) 14th
Free play	0.002 (0.001) 14th	0.002 (0.001) 13th	0.004 (0.002) 13th	0.005 (0.006) 13th
Curriculum links	0.002 (0.001) 15th	0.003 (0.001) 7th	0.003 (0.001) 15th	0.001 (0.001) 15th

Note. The app features are ordered based on the ranking for girls' literacy.

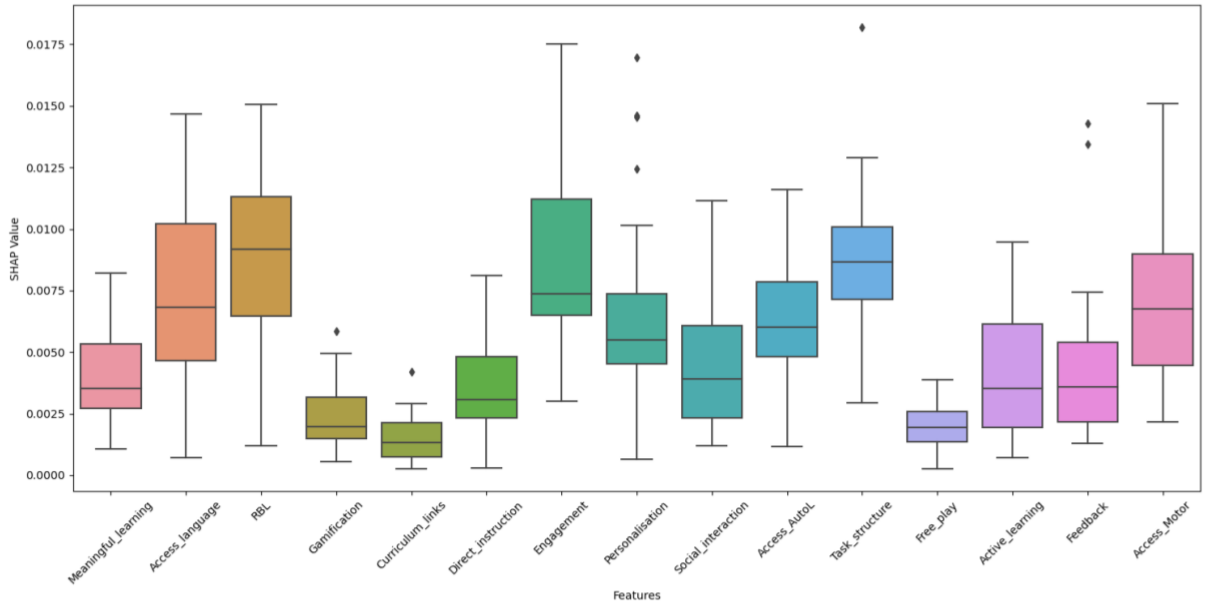
Looking across domains, engagement is arguably the most influential feature for both literacy and numeracy improvement, ranking highest for girls' numeracy and boys' literacy and second for girls' literacy and boys' numeracy. This suggests engagement is the most influential of the 15 app features investigated in predicting learning improvements. However, as shown in Figure 7, there is a large variability in SHAP values for girls' literacy, suggesting there may be interactions with other factors. Another impactful feature is retrieval-based learning, which ranked consistently high across all four categories and shows a more uniform influence for numeracy improvements (see Figure 7). Motor skills notably influenced boys' literacy and numeracy and were also influential for girls, scoring in the top six for all four categories, reflecting their broad relevance for learning. Task structure was similarly important across all four categories, especially in numeracy, where it ranked first for boys and second for girls. The box plots for motor skills and task structure highlight moderate variability in the SHAP values across domains and genders, suggesting that other factors may contribute to their impact on learning outcomes.

Interestingly, personalisation came sixth for girls' literacy, showing that it was moderately influential in predicting improvement in this domain. However, it showed little influence in other areas, ranking between tenth and fifteenth overall. Dispersion in Figure 7 demonstrates that the personalisation SHAP values are highly consistent and low for boys' literacy, showing little influence on improving boys' learning in this area. Gamification and free play showed limited influence across all domains and genders, consistently scoring in the bottom three app features for each category and scoring average SHAP values of 0.005 or less. Across all four categories, the SHAP scores for gamification and free play were consistently low, suggesting that as app features, they are not particularly influential for app-based foundational learning.

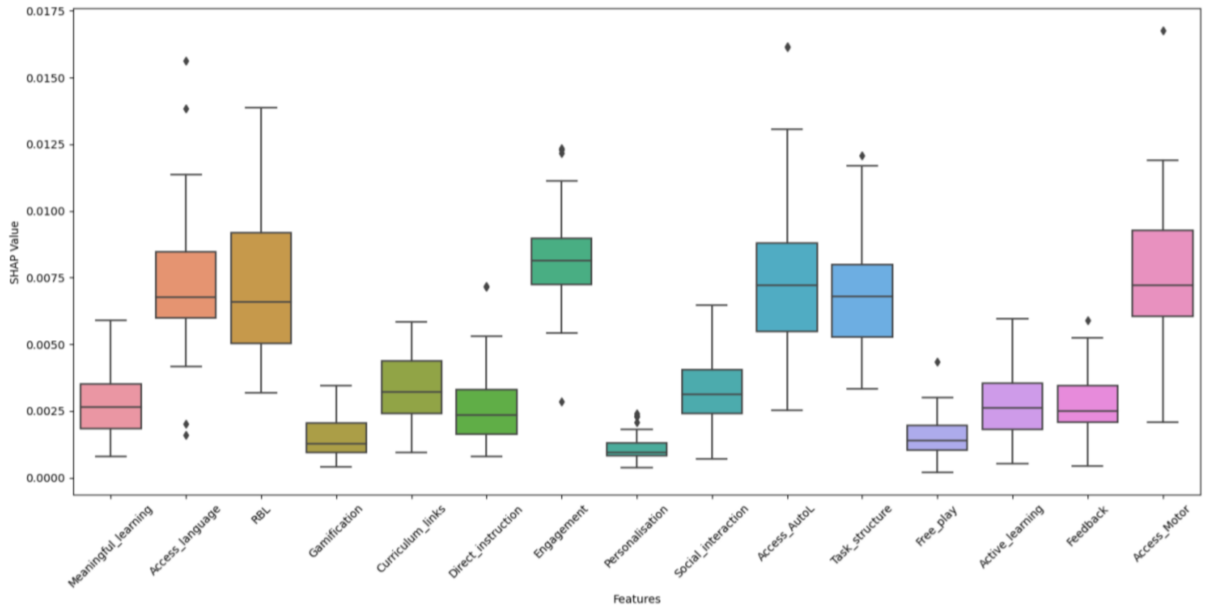
Figure 7

Box plots depicting the distribution of the 30 SHAP values for each gender per domain

Literacy

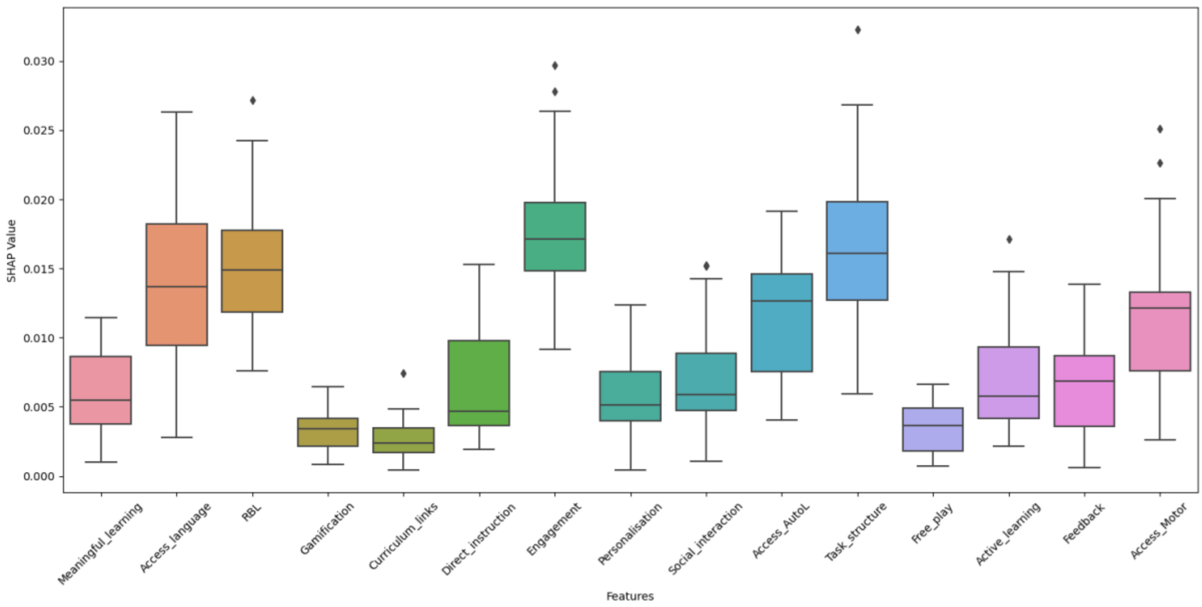


A. Girls' Literacy

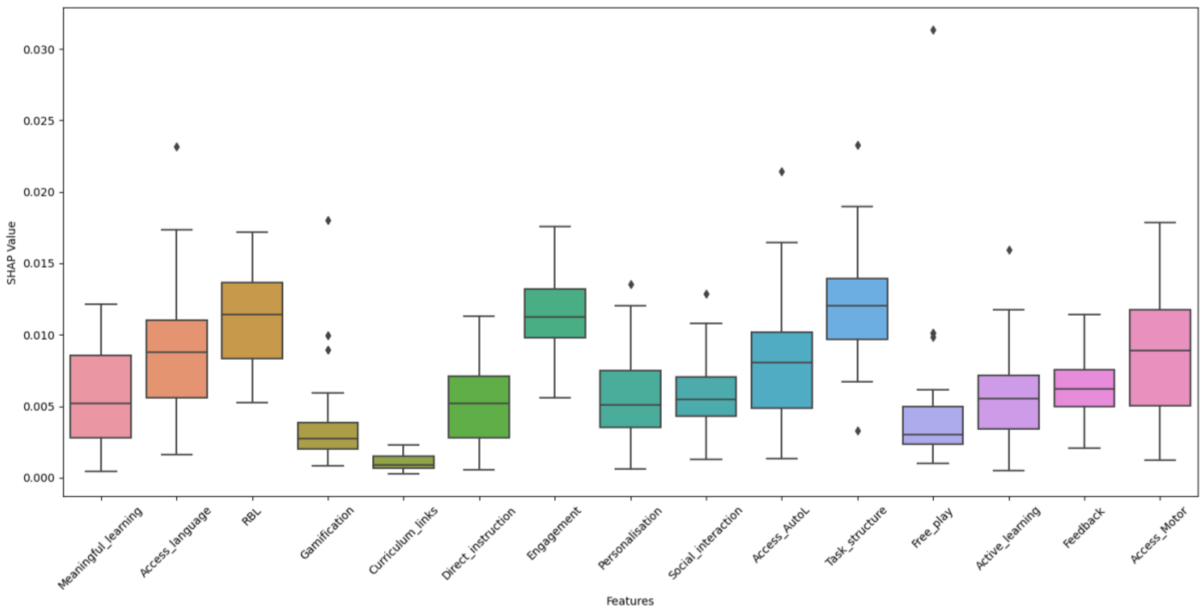


B. Boys' Literacy

Numeracy



C. Girls' Numeracy



D. Boys' Numeracy

Inferential tests

Given the insights provided by the SHAP analysis regarding the influence of each app feature, subsequent analyses utilising 2x2 ANOVAs and post-hoc t-tests were conducted to determine if there were any statistically significant differences for each app feature across domains and gender (see Table 10). The main effects for domain and gender are reported, but the interaction effects are the primary focus of the current research.

First, Shapiro-Wilks tests were conducted on the residuals from the ANOVA for each of the 15 app features to test whether they were normally distributed, using an alpha level of 0.05. It was found that 11 features significantly departed from normality ($p<0.05$). However, the residuals for four features - retrieval-based learning ($W = 0.99$, $p = 0.514$), motor skills ($W = 0.981$, $p = 0.08$), meaningful learning ($W = 0.982$, $p = 0.099$) and language demand ($W = 0.986$, $p = 0.228$) - exhibited normal distributions. Despite 11 tests violating the assumption of normality, as mentioned, ANOVA is considered a robust test, so the parametric ANOVAs were reported and are shown in Table 10. For the app features that violated assumptions of normality, the corresponding post-hoc tests were checked by conducting and reporting the equivalent non-parametric test where it differed from the parametric results. This was Mann-Whitney for the one sample t-tests (across gender for each domain) and Wilcoxon Signed Ranks for the paired comparisons t-tests (within gender across domains). The mean (SD) values for all app features are shown in Table 9 (above), and the ANOVA and post-hoc test results are reported in Table 10.

Table 10

Results of the 2x2 ANOVAs for each app feature: F-values, p-values, effect sizes and the corresponding post-hoc t-test comparisons for significant interactions

App feature	**Main effects**		**Domain* Gender interaction**	**Post-hoc tests**
	Domain F-value p-value η2	**Gender** F-value p-value η2	F-value p-value η2	
Curriculum links	6.815 0.010 0.055	0.427 0.515 0.004	65.430 <.001** 0.361	**Domain** Literacy vs. Numeracy (Girls): $t(29) = -3.618$, $p = 0.001$**, $d = -0.660$ Literacy vs. Numeracy (Boys): $t(29) = 9.413$, $p = <.001$**, $d = 1.719$ **Gender** Girls vs Boys (Literacy): $t(58) = -6.191$, $p = <.001$**, $d = -1.599$ Girls vs Boys (Numeracy): $t(58) = 5.250$, $p = <.001$**, $d = 1.355$

App feature	Main effects		Domain* Gender interaction	Post-hoc tests
	Domain F-value p-value η2	**Gender** F-value p-value η2	F-value p-value η2	
Personalisation	13.286 <.001** 0.103	28.421 <.001** 0.197	22.020 <.001** 0.160	**Domain** Literacy vs. Numeracy (Boys): t(29) = -7.233, *p* = <.001**, d = -1.321 **Gender** Girls vs Boys (Literacy): t(58) = 7.705, *p* = <.001**, d = 1.989
Engagement	80.500 <.001** 0.410	24.424 <.001** 0.174	15.106 <.001** 0.115	**Domain** Literacy vs. Numeracy (Girls): t(29) = -7.178, *p* = <.001**, d = -1.310 Literacy vs. Numeracy (Boys): t(29) = -5.759, *p* = <.001**, d = -1.051 **Gender** Girls vs Boys (Numeracy): t(58) = 5.535, *p* = <.001**, d = 1.429

App feature	Main effects		Domain* Gender interaction	Post-hoc tests
	Domain F-value p-value η2	**Gender** F-value p-value η2	F-value p-value η2	
Autonomous learning	20.493 <.001** 0.150	2.846 0.094 0.024	14.325 <.001** 0.110	**Domain** Literacy vs. Numeracy (Girls): t(29) = -6.643, *p* = <.001**, d = -1.213 **Gender** Girls vs Boys (Numeracy): t(58) = 3.325, *p* = 0.002**, d = 0.858
Language demand	26.800 <.001** 0.188	9.100 0.003** 0.073	9.154 0.003** 0.073	**Domain** Literacy vs. Numeracy (Girls): t(29) = -5.358, *p* = <.001**, d = -0.978 **Gender** Girls vs Boys (Numeracy): t(58) = 3.570, *p* = 0.001**, d = 0.922

App feature	Main effects		Domain* Gender interaction	Post-hoc tests
	Domain F-value p-value η2	Gender F-value p-value η2	F-value p-value η2	
Motor skills	10.778 0.001** 0.085	3.150 0.079 0.026	5.379 0.022 0.044	N/A
Retrieval-based learning	59.902 <.001** 0.341	18.083 <.001** 0.135	2.957 0.088 0.025	N/A
Task structure	70.617 <.001** 0.378	15.580 <.001** 0.118	2.456 0.120 0.021	N/A

App feature	Main effects		Domain* Gender interaction F-value p-value η2	Post-hoc tests
	Domain F-value p-value η2	**Gender** F-value p-value η2		
Gamification	18.774 <.001** 0.139	0.533 0.467 0.005	2.444 0.121 0.021	N/A
Free play	18.636 <.001** 0.138	0.501 0.481 0.004	2.153 0.145 0.018	N/A
Meaningful learning	29.987 <.001** 0.205	2.419 0.123 0.020	1.603 0.208 0.014	N/A

App feature	Main effects		Domain* Gender interaction	Post-hoc tests
	Domain F-value p-value η2	**Gender** F-value p-value η2	F-value p-value η2	
Feedback	35.484 <.001** 0.234	3.402 0.068 0.028	1.285 0.259 0.011	N/A
Direct instruction	30.974 <.001** 0.211	5.432 0.022 0.045	0.346 0.557 0.003	N/A
Social interaction	25.772 <.001** 0.182	5.127 0.025 0.042	0.046 0.830 <.001	N/A

App feature	Main effects		Domain* Gender interaction	Post-hoc tests
	Domain F-value p-value η2	**Gender** F-value p-value η2	F-value p-value η2	
Active learning	29.150 <.001** 0.201	7.100 0.009** 0.058	0.013 0.909 <.001	N/A

**p<.01. *Note.* The app features are ordered according to the magnitude of the effect size for the interaction, as this is the focus of the research.

ANOVA results

For the domain*gender interaction, curriculum links, personalisation, engagement, autonomous learning and language demand showed significance at the 0.01 level, highlighting the complex interplay between domain and gender for these app features. Again, effect sizes are small, with curriculum links having the highest effect size with d = 0.361. Post-hoc comparisons were undertaken using t-tests for the five significant app features.

Post-hoc comparisons

Results of the post-hoc tests determine where the significant differences lie within the significant interactions. See Table 9 for Mean (SD) scores for girls and boys for each literacy and numeracy improvements, for each of the 15 app features.

Significant differences were found between domains (literacy vs numeracy) for girls for four out of the five significant app features. There were large negative effect sizes for engagement (d = -1.310), autonomous learning (d = -1.213) and language demand (d = -0.978), and a moderate effect size for curriculum links (d = -0.660), indicating that these four app features positively influenced girls' numeracy achievement significantly more than literacy.

For boys' domains, significant differences were found for three app features: curriculum links, personalisation and engagement. Large negative effect sizes were found for personalisation (d = -1.321) and engagement (d = -1.051), implying that these app features positively influenced boys' numeracy achievement more than literacy. However, the opposite was found for curriculum links, which had a large positive effect size (d = 1.719), showing that curriculum links have a stronger positive influence on boys' literacy than numeracy achievement.

Significant differences were found in only two app features when comparing gender differences in literacy: curriculum links and personalisation. There was a large negative effect size for curriculum links (d = -1.599), demonstrating that curriculum links positively influence boys' literacy achievement more than girls. In contrast, personalisation had a large positive effect size (d = 1.989), indicating that personalisation favours girls' literacy achievement more than boys'.

When exploring gender differences in numeracy, significant differences were found for four of the five app features. There were large positive effect sizes for curriculum links (d = 1.355), engagement (d = 1.429), autonomous learning (d = 0.858) and language demand (d = 0.922), implying that these four features are more positively associated with girls' achievement in numeracy than boys.

Discussion

This study investigated which app features most effectively predicted literacy and numeracy improvement in Tanzanian out-of-school children, focusing on girls and boys independently, following an Edtech intervention. A multi-methodological approach was employed, combining the Bradley-Terry model, binary classification models, and 2 x 2 ANOVAs for each app feature with post-hoc t-tests. This allowed for an in-depth exploration of app feature importance across both genders and domains, including the significance of any observed differences found.

The Bradley-Terry analysis highlighted the XPRIZE winner onebillion as the most effective app across many of the 15 features, with personalisation being a unique strength of the other XPRIZE winner, Kitkit School. Overall, the winning apps had much higher lambda scores than the three apps that did not win the GLXP, indicating that they embodied the key app features more than the losing apps. This corroborates the results found in Chapter 3. The classification models showed a modest predictive ability for literacy and numeracy improvements following an app intervention, with the models predicting girls' performance more accurately than boys, albeit only marginally. This indicates that the chosen app features might be more influential generally for girls than boys, but not substantially so. This suggests that EdTech interventions targeted at autonomous learning of foundational skills might be more suited to girls than boys. The SHAP results suggested that over all app features, engagement was the most influential, ranking first or second for all four categories. Retrieval-based learning, motor skills, task structure and language demand were also highly influential app features, scoring in the top six features across all categories, reflecting their broad relevance for learning improvements. Accordingly, other learning apps that embody these pedagogical features should be effective at promoting foundational skills.

The ANOVA interaction effects highlighted complex relationships between gender and domain, with significant interactions found for five app features: curriculum links, personalisation, engagement, autonomous learning, and language demand. Post-hoc comparisons highlighted the nuanced differences across the four categories. On the domain level, app features such as curriculum links, engagement, autonomous learning, and language demand were better at predicting numeracy improvements than literacy for girls. For boys, engagement and personalisation were more predictive of numeracy improvement than literacy, while curriculum links favoured literacy over numeracy. Interestingly, autonomous learning and language demand did not significantly favour either domain for boys.

For gender comparisons, there were further distinctions; for literacy, personalisation favoured girls, while curriculum links were more beneficial for boys. No significant gender differences were found for literacy among engagement, autonomous learning, and language demand. In numeracy, significant differences were found for four app features: curriculum links, engagement, autonomous learning, and language demand, all of which were more influential for girls than boys. This analysis highlights the different influences that app features can have across domains and gender, emphasising the importance of tailoring app-based learning to the targeted domain and demographic to optimise learning improvements. As the literature looking at gender and domain differences for learning with apps is scarce, these findings add to the theoretical understanding of app-based learning.

This research suggests that engagement in the learning process (namely, whether in-app activities were appropriately engaging or disruptive and distracting) was arguably the most influential app feature for learning. This supports the findings in Chapter 3, which highlighted engagement as one of the top six app features across the two winning apps. The influence on numeracy across both genders highlights its broad relevance, indicating that this app feature would be highly influential for app-based numeracy learning. Previous research has also attributed engagement in the learning process as one of the key app features to support young children's learning (Hirsh-Pasek et al., 2015).

Retrieval-based learning was shown to be another key feature across all categories, consistently ranking high but particularly influential in improving girls' literacy. The broad effectiveness of retrieval-based learning suggests that embedding quizzes or tasks that test children's knowledge and understanding within an educational app can improve outcomes by reinforcing learning. Retrieval-based learning also featured highly in one of the two apps that were joint winners of the GLXP, onebillion, although it did not feature as one of the top six features across both winners in Chapter 4. The importance of retrieval-based learning in supporting literacy and numeracy acquisition aligns with previous research that has demonstrated the efficacy of retrieval practices in promoting meaningful long-term learning across multiple populations and contexts (Karpicke, 2017) and within computer-based learning for multiple domains (e.g. Science, Grimaldi & Karpicke 2014; and Numeracy, Pitchford, 2015).

Accessibility was proven to be critical for learning outcomes throughout this research, with all three accessibility-based app features - motor skills, autonomous learning, and language demand - found to be highly influential and ranked in the top seven app features across all four categories. This consolidates findings from Chapter 3 and previous research, which has emphasised the importance of these accessibility-based adjustments in literacy (e.g. Byrnes & Wasik, 2019) and numeracy (e.g. Outhwaite et al., 2020; Gulliford et al., 2021) app-based learning.

Personalisation was fairly influential for girls' literacy but had a variable impact across other areas, with a particularly low influence on boys' literacy. ANOVA results indicated that personalisation strategies should be tailored specifically to enhance girls' literacy and boys' numeracy outcomes, as this is where personalised levelling could have the most tangible positive impact. These findings add to previous literature that has shown personalisation with programmatic levelling to maximise learning outcomes in both literacy and numeracy app interventions (Vanbecelaere et al., 2020b; Outhwaite et al., 2023) by demonstrating how personalisation levelling is influenced by gender.

Curriculum links showed an interesting pattern of influence on learning outcomes, as they were moderately influential in predicting boys' literacy achievement but had a limited impact elsewhere. ANOVA results emphasised the stronger influence on boys' literacy over numeracy and showed a more pronounced effect on numeracy than literacy for girls, suggesting that a differentiated approach is needed to integrate curriculum links into educational apps to cater to the differing needs of girls and boys. The relatively limited influence of curriculum links resonates with findings in Chapter 4, where curriculum links featured significantly in only one of the two winning apps, suggesting that links to the curriculum are not essential for learning foundational skills with an educational app.

For gamification, it was anticipated that this would have more of an influence on learning for boys than girls, as suggested by previous literature citing a higher level of competition and desire for rewards within an app (Kickmeier-Rust et al., 2014; Yang & Quadir, 2018). However, results showed that gamification had very little influence across the four categories, indicating that its impact on learning outcomes for both genders may be limited.

Novelty of research methods

A core strength of this research is the novelty of the methodological approach, combining machine learning methods with inferential statistics. While big data has begun to be utilised in educational research (Yu, 2020), the standard approach to running a classification model followed by SHAP analysis would be to run the classification model multiple times before selecting the best-performing model to run a SHAP analysis and get a singular set of SHAP values. To enable the exploration of gender and domain effects on app features, this study ran the SHAP analysis multiple times and averaged the resulting SHAP values. This innovative approach enabled inferential statistics to determine whether differences between SHAP values were significant, allowing comparisons across and within domains and genders. This combination leverages the strengths of both individual analytical approaches; the machine learning approach handles large, complex datasets well and allows the identification of the key predictors of learning outcomes. However, prediction alone is not enough; an explanation is also needed to understand the causal pathways and where differences lie between domains and gender (Lavelle-Hill, 2020). Using inferential statistics allowed a further understanding of the significant relationships, which, in combination with the predictive accuracy, can produce a more informed, nuanced understanding of how app features influence learning outcomes differently depending on the domain or gender being targeted.

Limitations

The primary limitation of this study is the model's performance in these binary classification tasks, which was shown to be only slightly better than chance for both EGRA and EGMA improvement scores. This may reflect the use of lambda scores in the models that were provided by naive participants who were unfamiliar with the context in which the GLXP took place. It may also be attributable to the complexity of the learning process, which is likely to be influenced by a range of factors, including individual differences such as personal experiences, prior knowledge, and diverse learning styles (Jonassen & Grabowski, 2012; Cuthbert, 2005). Factors pertaining to the environment in which the child resides and factors attributable to the child and government education policy may also influence learning outcomes with educational apps, as Pitchford (2023) discussed. The effectiveness of app-based learning in LMIC remote community-based settings is also likely influenced by how educational apps are implemented, as implementation has been shown to be critical in determining learning gains with educational apps in school-based settings (e.g., Outhwaite et al., 2019b). These are critical factors that the models reported here do not consider, yet they are likely play a significant role in affecting learning outcomes. For a comprehensive understanding of how children learn foundational skills with educational apps, research is needed to investigate each of these potential influences and then synthesise findings across studies. This will be explored in Chapters 5 (Contextual Predictors) and 6 (Expert Elicitation).

Another potential limitation of this research is setting the significance level at 0.01 in an attempt to balance the risk of Type 1 and Type 2 errors. Although it is less conservative than a Bonferroni-adjusted level, the chosen significance level may still carry a risk of Type 2 errors, leading to the possibility of overlooking true effects. While it reduces the likelihood of false positives, 0.01 may not be optimal for detecting subtler effects, especially in a study with multiple comparisons. Choosing a 0.01 threshold may also affect the comparability of the study's results with other research in the EdTech field, where significance levels of 0.05 are often adopted (e.g. Walter-Laager et al., 2017; Fadhli et al., 2020).

Finally, this study did not look at how different app features interact and how effective they are in combination. Certain combinations of app features may be more effective for girls or boys in enhancing foundational outcomes. For example, personalised levelling combined with feedback has been found to be influential for learning maths with educational apps (Outhwaite et al., 2023). Understanding how different app features interact could contribute towards the design of more gender-responsive educational apps.

Conclusion

This study adopted a novel and innovative analytical approach to identify which app features are most influential for learning improvements in literacy and numeracy for out-of-school boys and girls in Tanzania, by exploring differences across domains and genders. Some app features, such as engagement, were found to be broadly influential regardless of the domain or gender, while others, such as personalisation, were more successful for a targeted audience, specifically out-of-school girls using apps to learn literacy skills.

While many app features showed a similar influence across literacy and numeracy for boys and girls, five app features showed a significant interaction demonstrating these app features exerted a differential influence across genders and domains. These app features were curriculum links, personalisation, engagement, autonomous learning and language demand. These novel results should inform the future pedagogical design of educational apps, as these findings suggest that app design can be tailored to enhance learning outcomes in particular domains differentially for boys and girls. This could be particularly effective in the global plight to improve access to quality education and reduce the attainment gap for girls (UK Aid, 2019).

Chapter 5: Contextual Predictors of Foundational Learning Outcomes

Background

Chapters 3 and 4 have explored the role that different app features play in facilitating learning outcomes with out-of-school children following an EdTech intervention, identifying which features were most influential for foundational literacy and numeracy improvements, and exploring differences based on specific domain and genders. While informative for app design, most EdTech initiatives do not take the local context into consideration when designing and implementing learning app software with out-of-school children. Poorly implemented interventions risk the exacerbation of existing learning inequalities, highlighting the need for an informed contextual understanding (Tauson & Stannard, 2018; Allison, 2023). Furthermore, app features alone cannot account for the variability in performance shown by individual children that participated in the GLXP, as some children made greater improvements with this educational technology than others, even whilst receiving instruction with the same app.

Contextual factors are suggested to play a vital role in the global learning crisis (e.g. Huntington et al., 2023a; Zubairi et al., 2021), yet have received little attention to date due to an insurmountable deficit in traditional census and survey data that can be linked directly to learning outcomes. To address the limitations highlighted in the previous chapters and begin to build the evidence base around this low-income context, the current study explored potential child-level and village-level contextual predictors to identify which features are most predictive of children's learning improvements following the GLXP EdTech intervention. Understanding factors associated with positive learning outcomes in remote settings with out-of-school children is vital to enhance theoretical understanding and promote effective interventions.

Previous research has established several child-level contextual factors that have been considered strong predictors of academic success, including demographic characteristics such as gender (UNESCO, 2015; OECD, 2016), and educational characteristics, such as school attendance (Gulliford & Miller, 2023) or literacy levels of caregivers, siblings and other family members in the household (e.g. Anders et al., 2012). In addition, contextual factors at the village level, including community characteristics, such as access to healthcare (Porter, 2014), availability of financial or educational resources (Ayiro & Sang, 2016), and physical infrastructure, such as transportation and road networks (Huntington et al., 2023a), may also influence children's learning outcomes in LMICs.

Existing reviews have explored what works to improve learning outcomes in developing countries (Evans & Popova, 2016; McEwan, 2015), including Sub-Saharan Africa (Conn, 2017). Factors attributed to improvements in foundational learning include teacher-specific variables, such as training and subject knowledge, and organisational variables, such as students having their own learning space and smaller class sizes that are grouped by ability. While valuable, this research was conducted with children in formal schooling, so most of the significant variables refer specifically to a school context. Much less is known about contextual factors that influence learning outcomes for out-of-school children, yet this population must be investigated further, as they are among society's most vulnerable and marginalised members, with the out-of-school population in Sub-Saharan Africa continuing to grow (UNESCO, 2022c). These children need effective interventions to acquire core foundational skills for future health, wealth, and well-being.

The current study addressed this gap in knowledge by exploring predictors of foundational learning outcomes for out-of-school children in Tanzania who used interactive apps delivered on tablet-technology as part of the GLXP competition. Identifying factors that predict learning outcomes for out-of-school children with and without access to educational technology is necessary to enhance understanding of how to design effective interventions to improve educational outcomes for this hard-to-reach, marginalised population.

Current Study

This study aimed to identify contextual factors that are predictors of, and potentially facilitators for, successful learning of foundational skills by out-of-school children in LMICs with and without educational technologies. The following research questions were investigated:

> RQ1. Which factors best predict foundational learning skills in out-of-school children in remote areas of Tanzania?
>
> RQ2. Which factors best predict improvements in foundational learning outcomes after implementing an educational technology intervention with out-of-school children in remote areas of Tanzania?

Given the complexity of potential predictors of learning outcomes, a machine learning approach is needed to investigate these questions, as demonstrated in Chapter 4, due to the ability to overcome challenges of traditional data analysis, such as overfitting and multicollinearity (Lavelle-Hill et al., 2021). Accordingly, the current study evaluated different data analytical methods to address the two research questions, namely Linear Regression - a traditional data analysis method, compared to Decision Trees, Support Vector Regression (SVR), Random Forest (Regressor), and XGBoost - all machine learning data analytic methods. A series of 12 individual regressor prediction experiments were conducted to determine (i) child features only, (ii) village features only, and (ii) child and village features combined that best predicted learning outcomes for (a) literacy (EGRA scores) and for (b) numeracy (EGMA scores), before (RQ1) and after (RQ2) intervention with the educational technologies implemented in the GLXP. The highest-performing model was used for each experiment to identify the strongest predictors of foundational learning outcomes.

Child features

In the current study, child features were considered any personal, demographic, or environmental characteristics related to the child that may impact learning outcomes. Potential predictors were chosen from those suggested by the extant literature to influence children's learning outcomes. The list of child features used in the current study was not exhaustive, as data available for the sample population was limited to that collected during the GLXP competition.

Demographic variables that affect children's cognitive and social development were included, such as age and gender. Engagement levels and attainment with learning programmes may differ between gender, due to the previously mentioned competing demands with household chores (Miheretu, 2019).

Prior knowledge has also been shown to have a significant and positive impact on knowledge acquisition of learners in LMICs and has been perceived to support early maths learning and engagement when using educational technology in the UK (Tavera & Casinillo, 2020; Gulliford & Miller, 2023). This suggests that previous experiences with schooling may influence continued learning of basic literacy and numeracy skills when children are out of school. Therefore, school attendance before the GLXP field trial was included as a potential predictor for learning outcomes.

Other potential predictors were based on characteristics unique to the child's home environment. Children's early experiences with literacy are highly predictive of early reading ability and scholastic success, and levels of parental involvement in the home environment have been shown to correlate with children's literacy learning (Timmons & Pelletier, 2015; National Early Literacy Panel, 2008). Parental involvement can take many forms, such as engaging children with shared reading, writing activities, singing, and vocabulary games. Shared reading between parent and child has been shown to be an essential home learning activity for academic achievement (Timmons & Pelletier, 2015; Jeynes, 2012), and shared reading experiences during early childhood have been attributed to higher reading achievements, language abilities, and reaching of developmental milestones later in childhood (Boonk et al., 2018; Turesky et al., 2022). While the quantity of shared reading experiences has been shown to predict early reading achievement, mathematic achievement seemingly depends on the quality of discussion after reading a text (Barnes & Puccioni, 2017). Furthermore, caregiver involvement has been shown to facilitate the success of digital learning solutions in LICs (Islam et al., 2022), indicating that familial involvement could be pertinent in this context.

However, the link between parental socio-economic status (SES) and level of parental involvement is inconsistent. Barnes and Puccioni (2017) found that children from higher SES homes in the US were more likely to experience shared book reading with parents than children from lower SES homes, but these results could not be replicated in China (Cheung et al., 2022). The extent of parental involvement can be influenced by maternal education, with highly educated mothers showing a positive association with high levels of paternal involvement in both Spain and the UK. Additionally, there is an observed increase in maternal involvement in Spain when mothers are highly educated (Gimenez-Nadal & Molina, 2012). However, the correlation between a home literacy environment and reading ability may not arise primarily from shared reading experiences and book access. Instead, it is more likely attributed to heritability, which accounts for a genetic predisposition for reading abilities (van Bergen et al., 2017). Consequently, caregivers' literacy level, type of work, and home reading habits were included as potential predictors in the current study to further investigate the relationship between home literacy, SES, and learning outcomes of out-of-school children, with and without EdTech interventions.

These features reflect the child's prior experiences and exposure to learning opportunities, which may influence their acquisition of foundational skills. During the GLXP, data was collected for out-of-school children in Tanzania on foundational literacy and numeracy skills, at baseline, prior to intervention, and at endline, after the intervention, with educational technologies (as explained in Chapter 2). This data was used to identify the child features most strongly associated with foundational learning at baseline (RQ1) and improved literacy and numeracy skills after implementing the digital interventions deployed in the Global Learning competition (RQ2).

Village features (also referred to as community-level factors)

As the GLXP was conducted in 172 remote villages in Tanzania, village features that provide data on community and physical characteristics that may influence a child's ability to learn were chosen for investigation. Non-income poverty encompasses physical and social isolation caused by living in remote locations where access to goods and services is limited (Mbilu, 2019). Non-income poverty can impact children's learning outcomes significantly and become a vicious cycle whereby illiteracy and innumeracy further exacerbate social and physical isolation. While physical isolation is understandably most difficult for out-of-school children, local schools are also affected by the precarious infrastructure of more rural areas, as they struggle to recruit high-quality teachers, have difficulty accessing sufficient supplies and curriculum materials, and are negatively impacted by severe weather conditions (White, 2015).

Physical isolation via inaccessible environments has been identified as a significant constraint to participation in formal and non-formal education across different areas of low-income Africa (Ogbonna, 2015; Ayiro & Sang, 2016), and roads are a vital infrastructure that influences transportation and access to resources. Hence, in the current study, village accessibility was included as a potential predictor of learning outcomes, as measured by the number of roads in and out of the village and the quality of the roads.

High levels of transport poverty across remote areas of Africa further augment the impact of physical isolation. This results in overwhelming numbers of households not having regular access to any motorised transport, which undermines their ability to access critical economic and social activities and puts individuals at a significant disadvantage in accessing primary education (Lucas, 2011; Lemon & Battersby-Lennard, 2009). Good transport networks facilitate physical access to education and health services, which promotes community well-being and secures a healthy workforce that is pivotal for a country's economic growth (Porter, 2014). Hence, in the current study, distances to the nearest bus, train, and airports were used as village features to reflect the extent of isolation from the closest transport links.

Physical barriers can also negatively contribute to feelings of social isolation that are associated with poor developmental and educational outcomes (Leigh-Hunt et al., 2017). Social connectedness positively impacts mental and physical health (Diendorfer et al., 2021), so lacking a sense of belonging can contribute to feelings of hopelessness around accessing education, particularly for marginalised girls (Oulo, 2021). In this study, social connectedness was measured by (i) the number of children and buildings in the village as an indirect representation of the level of peer networks or community support that a child is likely to receive and (ii) the distance to vital local amenities and services, such as financial services, police stations, health centres, and places of worship.

Technology-based distance learning may mediate physical isolation by facilitating social connections in rural areas and providing new educational opportunities for marginalised children, which supports their psychosocial well-being (Ashlee et al., 2020; Dryden-Peterson et al., 2017). However, remote learning presents its own challenges, as access to educational technologies is often not equitable for girls, even when technology is available within the household (Amenya et al., 2021). Poor digital infrastructure, such as limited access to grid electricity and poor internet connectivity, can undermine the potential of enhanced education provided by technology (Ashlee et al., 2020; Huntington et al., 2023a).

Collectively, these village features reflect the child's physical and social environment, which may influence their access to education and accompanying resources that support learning. This geospatial data was collected online for the villages in Tanzania that took part in the GLXP and was used to identify the village features that were most strongly associated with foundational learning at baseline (RQ1) and improved literacy and numeracy skills after implementation of the digital interventions deployed in the Global Learning competition (RQ2).

Child and village features combined

To investigate how child and village features interact in predicting children's learning outcomes, a set of regressor prediction experiments was conducted that analysed data collected about the children who participated in the GLXP alongside geospatial data collected online about the corresponding Tanzanian villages. These experiments were conducted to identify which combination of child and village features were most strongly associated with foundational learning at baseline (RQ1) and which feature combinations were most strongly associated with improved literacy and numeracy skills after the implementation of the digital interventions deployed in the Global Learning competition (RQ2).

Method

Design

This study leverages two empirical datasets, one acting at the child-level and one at the village level, to address the two research questions outlined below by undertaking 12 individual regressor prediction experiments.

As we were interested in ascertaining factors that predict literacy and numeracy acquisition at baseline, prior to intervention, and at endline after implementation of an educational technology intervention, both EGRA and EGMA scores are used, respectively, as independent outcome variables across the analyses (see Chapter 2 for more details on the dependent variables used in the analyses).

In order to discern the relative importance of both child and village-level features a three-stage approach was used, first modelling base-level attainment and then improvements in literacy due to interventions solely through the lens of child-level features. Experiments were then repeated with village-level contextual features, and finally a combination of all features (resulting in 6 modelling experiments). This whole process was repeated for numeracy, resulting in a final methodology that consists of 12 experiments, allowing for examination of informational gain made available by each variable set. The overall study design is illustrated in Figure 8.

Figure 8

Mapping the 12 predictor tasks to literacy and numeracy baseline and improvement scores for child and village features.

The data

Child-level data

The child-level data used for the analysis in this chapter is taken from the primary XPRIZE dataset detailed in Chapter 2, consisting of the localised literacy and numeracy (EGRA and EGMA) assessments and a contextual survey about the child's home life and environment, conducted with the child and their caregiver and administered both at baseline and endline. The EGRA and EGMA raw assessment scores were used as the dependent variables for the machine learning models (see further details below). Relevant questions from the contextual survey were selected as predictor variables in the child-level models; these were chosen based on the features that research has suggested may impact learning outcomes, such as caregiver profession, prior experiences with schooling and technology, and number of siblings (as explored in the introduction).

Village-level data

XPRIZE data. The XPRIZE Foundation provided location (latitude / longitude) data of the 172 Tanga villages that participated in the competition for this research. XPRIZE also provided categorisation data about the 'reachability' of each village, using a 4-point scale ranging from easy to very hard. This was based on subjective judgments made by the XPRIZE implementation team working on the ground regarding ease of access to each participating village and considering both rurality and road quality. For example, a village near a large town with good quality roads accessible by car would be categorised as 'easy'. In contrast, a village in a highly remote location that could be accessed only on foot and with difficulty would be categorised as 'very hard'.

Geospatial data. For each participating village, geospatial data was collected for this research from open-source shapefiles containing information about the geographical locations of buildings, transport systems, local services, and amenities across Tanzania. Building data was obtained from the 'Building Footprints' initiative in which Bing Maps and Microsoft Philanthropies partnered with the Humanitarian OpenStreetMap Team (HOT) to use AI-assisted mapping to identify map features at scale, including a country-wide dataset of all buildings in Tanzania (Microsoft Bing, 2020). All other sources were open-source shapefiles found through online repositories and were created by HOT, last updated in April 2020, a year after completion of the endline data collection in the GLXP (Humanitarian Data Exchange, 2020). The collected geospatial data allowed mapping of the connectivity and isolation of the participating villages.

Feature Engineering

Before the datasets could be utilised in the regression analyses, the data was cleaned, and features engineered as outlined below.

Dependent variables

The first aim of this work was to determine features that best predict foundational learning skills for out-of-school children in Tanzania, so raw baseline scores for EGRA (/261) and EGMA (/101) were used as dependent variables in the literacy and numeracy analyses, respectively (RQ1).

The second aim of this work was to identify features that best predict improvements in learning outcomes following the implementation of an EdTech intervention with one of the five finalised apps compared to the control who did not receive an app-based intervention. For these analyses, the dependent variables were difference scores for literacy and numeracy skills over time, calculated by subtracting endline scores from baseline scores for each participating child for EGRA and EGMA, respectively (RQ2).

As our intended purpose was not to simply predict a child's learning score but to determine variables that could predict learning progress, the baseline learning scores were not used as a predictor for the improvement scores (RQ2). This also explains why a change score (improvement) was used rather than the raw endline scores.

Independent variables

As potential predictors of children's learning, 33 candidate features were engineered across the datasets described above. Table 11 reports a complete list of the features, a description of each feature, when the feature was used and where the data was collected from.

Relevant child-level features were extracted from baseline and endline survey datasets. Categorical variables were one-hot encoded during data pre-processing, as it is a simple and widely used encoding method (Cerda et al., 2018). One-hot encoded variables were 'caregiver work', 'village reachability' and 'type of app'. Median values were used to impute missing data, as 41 'village children' values and 441 ages were missing. Child-specific input features were engineered to preserve privacy by rigorously anonymising or removing all identifying information.

The village locations were aggregated with a 1km circle buffer around each to calculate reliable distances from each village to services. Features about the villages and their surrounding areas were extracted from open-source data. Building and road location data was used to engineer features. All buildings and roads within the 1km area were counted to create two measures, 'village roads' and 'village buildings', that can reflect the level of social connectedness a child may experience and the development status of their village. Input features from the remaining geospatial data (e.g. locations of the nearest transport, health, and financial services) were engineered to reflect physical isolation and connectedness by calculating the distance from the nearest transport, amenity, or service to each buffered village perimeter using SQL commands in PostGIS (e.g. Village 107 may be 5.4km from the nearest airport). Examples of these features include the distance to the nearest bank, the distance to the nearest train station and the distance to the nearest pharmacy. The geospatial data was then matched to each child from the GLXP based on which village they lived in.

Table 11

A complete list of the child and village-level features generated and their data source

Feature *As named in the analysis*	Data source	Used for: Baseline (B)/ Improvement (I) models	Description
Child-level features			
age	XPRIZE Survey	B, I	The reported age of the child
female	XPRIZE Survey	B, I	The reported gender of the child
often_child_reads	XPRIZE Survey	B, I	How often the child reads out loud to someone at home
often_child_read_to	XPRIZE Survey	B, I	How often the child is read to at home
used_tablet_before	XPRIZE Survey	B, I	Whether the child has ever used a device like a tablet before
attended_school_ before	XPRIZE Survey	B, I	Whether the child has ever attended school before
amount_siblings	XPRIZE Survey	B, I	The amount of siblings the child has (as reported by CG)
older_siblings_literate	XPRIZE Survey	B, I	The number of older siblings that can read and write – measured as all/some/none (as reported by CG)
younger_siblings_ literacy	XPRIZE Survey	B, I	The number of older siblings that can read and write – measured as all/some/none (as reported by CG)

Feature *As named in the analysis*	**Data source**	**Used for: Baseline (B)/ Improvement (I) models**	**Description**
caregiver_highest_ schooling	XPRIZE Survey	B, I	Highest level of schooling of the primary caregiver (as reported by CG)
caregiver_work	XPRIZE Survey	B, I	Job sector of the primary caregiver (as reported by CG)
CG_used_similar_ device_before	XPRIZE Survey	B, I	Whether the caregiver has used a similar device to the tablet before (as reported by CG)
caregiver_phone_ type	XPRIZE Survey	B, I	What type of phone the caregiver uses (as reported by CG)
electricity	XPRIZE Survey	B, I	Whether the child's home has electricity (as reported by CG)
home_radio	XPRIZE Survey	B, I	Whether the child's home has radio (as reported by CG)
home_mobile	XPRIZE Survey	B, I	Whether the child's home has a mobile phone (as reported by CG)
home_tv	XPRIZE Survey	B, I	Whether the child's home has a TV (as reported by CG)
eaten_todayBL	XPRIZE Survey	B	Whether the child has eaten on the day of assessment (at baseline; this information was not collected at endline)
village_mama_help	XPRIZE Survey	I	Whether the village mama helped the child with the tablet throughout the intervention (Yes/No)
household_adult_ help	XPRIZE Survey	I	Whether a household adult helped the child with the tablet throughout the intervention (Yes/No)

Feature *As named in the analysis*	Data source	Used for: Baseline (B)/ Improvement (I) models	Description
sibling_help	XPRIZE Survey	I	Whether a sibling helped the child with the tablet throughout the intervention (Yes/No)
Village-level features			
village_children	XPRIZE	B, I	Number of children in each village (as reported by XPRIZE facilitators conducting the survey)
village_buildings	Geospatial – Bing BF	B, I	Number of buildings in each village (within 1KM buffer)
village_roads	Geospatial - HOT	B, I	Number of roads in each village (within 1KM buffer
waterway_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest waterway in KM
railway_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest railway in KM
health_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest pharmacy, doctors, health centre or hospital in KM
finance_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest ATM or bank in KM

Feature *As named in the analysis*	Data source	Used for: Baseline (B)/ Improvement (I) models	Description
placeofworship_ distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest place of worship in KM
police_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest police station in KM
busstation_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest bus station in KM
airport_distance	Geospatial - HOT	B, I	Distance from the village (1KM buffer) to the nearest local airport in KM
village_reachability	XPRIZE	B, I	Measure of road quality and rurality based on judgements from the XPRIZE implementation team

Note. CG = caregiver

Modelling Approach

A core regression task was formulated to investigate the importance of factors that may predict children's learning outcomes in rural Tanzania. Machine learning models were established based on the sample of 2041 children in which each child was described by their child and village-level factors. Competing regression models were evaluated, including a 'Dummy Regressor' (as a baseline comparison for the other models), Linear Regression, Decision Trees, Support Vector Regression (SVR), Random Forest (regressor) and XGBoost. These predictive models were chosen as they produced interpretable outputs for variable importance, which was the primary focus of the analysis.

As in Chapter 4, the data was split into a training set (80%; N = 1633) and a test set (20%; N= 408). The performance of each model was assessed using a K-fold cross-validation approach where K=5. For each model, grid searches were performed using the 80% training subsample to identify optimal hyperparameters. The predictive performance of each model class was then measured on the held-out test set using the optimised parameters to prevent overfitting, evaluated by R-squared scores (R2; McFadden, 1973), mean absolute error (MAE) and mean squared error (MSE) values. This process was repeated 30 times, giving 30 scores for the final model performance.

The best-performing model class (identified using the highest mean R2 score and lowest error scores) was then chosen, which for all 12 analyses was either Random Forest (five analyses) or XGBoost (seven analyses). As previously established, Random Forest is an ensemble tree-based method that is very efficient with large databases and more robust to errors and outliers than other boosting methods (Han et al., 2022). XGBoost, short for eXtreme Gradient Boosting, is a highly efficient optimised gradient-boosted decision tree system that provides parallel tree boosting (Chen & Guestrin, 2016).

The Random Forest and XGBoost models were fitted to the complete datasets with the optimal parameters to identify the most important predictive features. The R2 scores were checked to ensure no significant decrease in performance, and then analyses of variable importance were undertaken using SHAP to understand how different feature values affected the literacy and numeracy outcome measures (Lundberg & Lee, 2017).

Results

For each of the 12 experiments, results for the best-performing models are reported followed by model interpretation using SHAP scores to address each of the two research questions posed.

RQ1: Which factors best predict foundational learning skills in out-of-school children in remote areas of Tanzania?

Model performance

Table 12 reports results for the six experiments conducted to answer the first research question. For each experiment, the best-performing model class is reported along with the model's average performance over 30 iterations on the held-out test set. As can be seen, R2 scores were low (between 0.08 and 0.22) for all experiments, but the best-performing model showed an overall improvement to the dummy model and thus allowed for the most accurate prediction of children's learning scores. Comparing model performance scores across the six experiments, child-level factors were most predictive of children's baseline learning outcomes, and model performance improved only marginally when child and village factors were combined.

Table 12

Summary of the performance of the six regressor models conducted to determine factors that best predict foundational learning skills in out-of-school children in remote areas of Tanzania (RQ1).

Dependent Variable	**Feature Set / Experiment**	**Dummy Performance**	**Best Performing Model Class**	**Model Performance**
Literacy (baseline)	Child (Exp 1)	R2: 0.00 MAE: 7.05 MSE: 219.43	Random Forest	R2: 0.16 MAE: 6.31 MSE: 165.86
	Village (Exp 2)	R2: 0.00 MAE: 6.89 MSE: 204.63	XGBoost	R2: 0.08 MAE: 6.24 MSE: 186.16
	Combined Child + Village (Exp 3)	R2: 0.00 MAE: 6.99 MSE: 210.89	Random Forest	R2: 0.18 MAE: 6.24 MSE: 180.27
Numeracy (baseline)	Child (Exp 4)	R2: 0.00 MAE: 7.36 MSE: 112.47	XGBoost	R2: 0.21 MAE: 6.22 MSE: 87.88
	Village (Exp 5)	R2: 0.00 MAE: 7.34 MSE: 111.66	XGBoost	R2: 0.09 MAE: 6.77 MSE: 103.62

Dependent Variable	Feature Set / Experiment	Dummy Performance	Best Performing Model Class	Model Performance
	Combined Child + Village (Exp 6)	R2: 0.00 MAE: 7.46 MSE: 117.65	XGBoost	R2: 0.22 MAE: 6.13 MSE: 85.96

Variable importance

As the aim of this research was to determine which variables best predict foundational learning outcomes, the best-performing model class with optimised hyperparameters were fitted to the entire dataset to investigate variable importance. SHAP values were then calculated, allowing for interpretation of the direction of the relationship between each predictor and literacy and numeracy outcomes, along with the degree to which each variable contributed to the overall model. Figure 9 reports SHAP summary plots for each of the six experiments conducted to address RQ1.

At the child level, the most influential factors predicting successful learning outcomes for literacy and numeracy were prior school attendance, how often the child read or are read to, and age. Age was particularly important for numeracy, as it had the greatest variability within the model, and revealed that high ages corresponded to a positive impact on the model. Having few siblings also positively impacted the model, albeit with a less pronounced clarity, as the more blended distribution of SHAP values indicates a less direct relationship with the numeracy outcome scores compared to age. Results also showed that high exposure to technology and high household literacy had a positive but marginal impact on the prediction of learning outcomes. Somewhat surprisingly, unemployed caregivers had a higher positive impact on the model than any given profession.

Relationships were less clear at the village level than at the child level. Results at the village level showed the most influential predictors of literacy and numeracy outcomes were proximity to a police station and the number of roads in a village, both of which showed positive impacts on the prediction model. Distance to health services had a negative relationship with the outcome variables, demonstrating that lower levels of foundational learning were found in villages that were furthest away from health services. In addition, the number of children in a village was shown to have a negative marginal impact on predicting numeracy learning outcomes, and reachability of a village had the lowest overall impact on model predictions compared to the other village-level factors investigated.

Figure 9

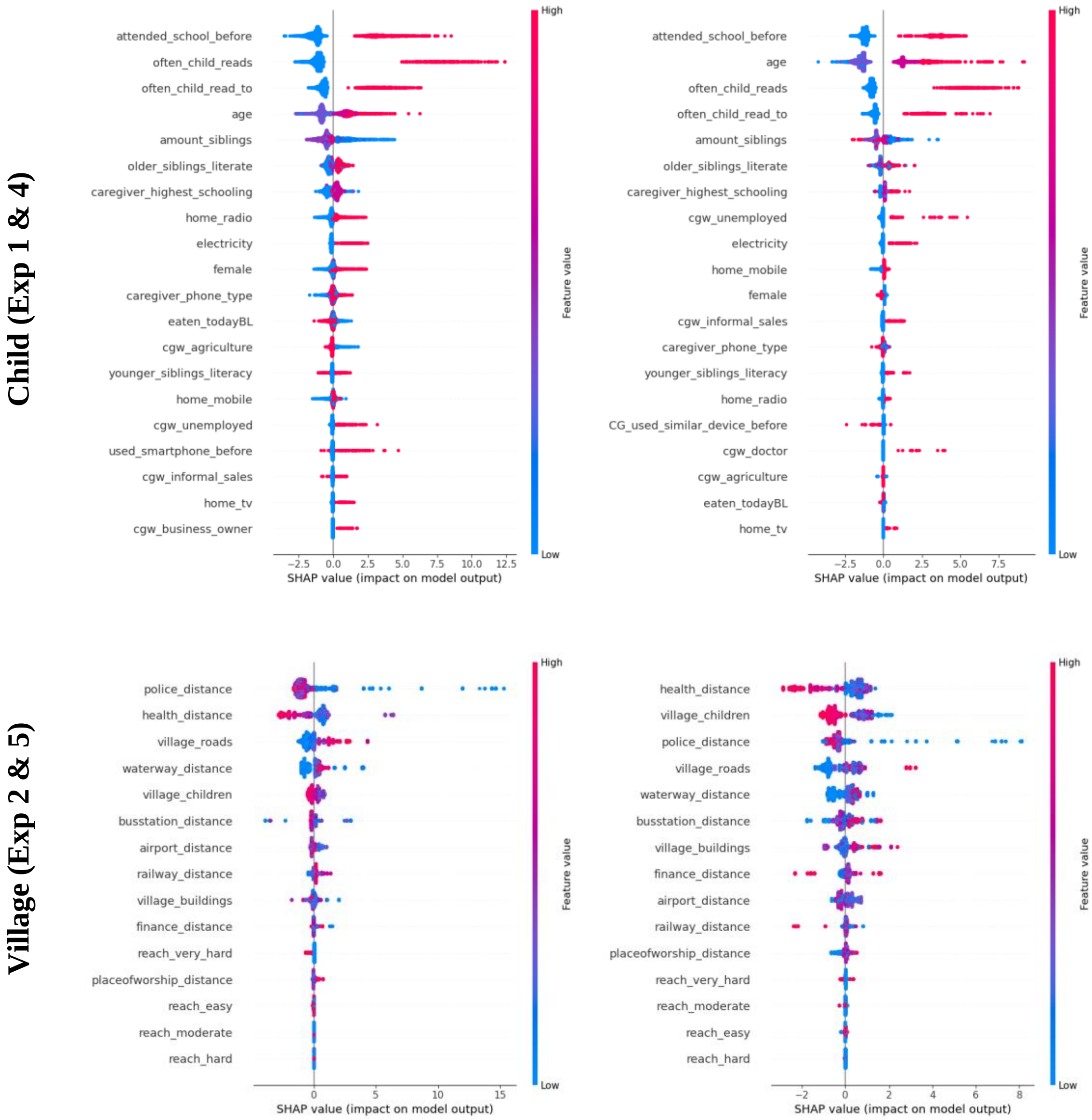
SHAP summary plots for the six experiments conducted to determine factors that best predict basic literacy and basic numeracy skills in out-of-school children in remote areas of Tanzania (RQ1)

Outcome: Literacy

(EGRA baseline scores)

Outcome: Numeracy

(EGMA baseline scores)



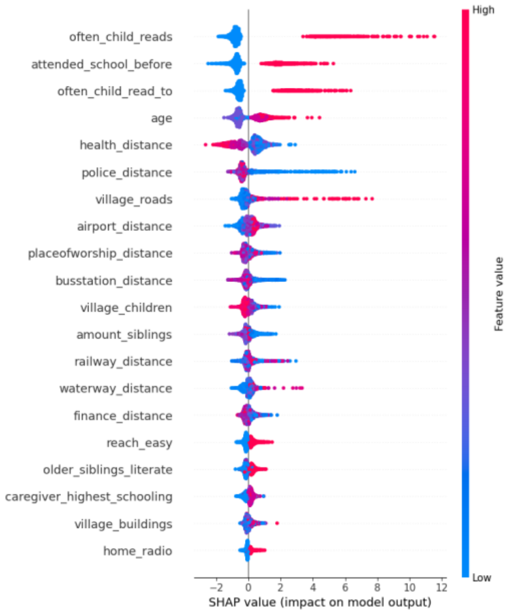
Outcome: Literacy

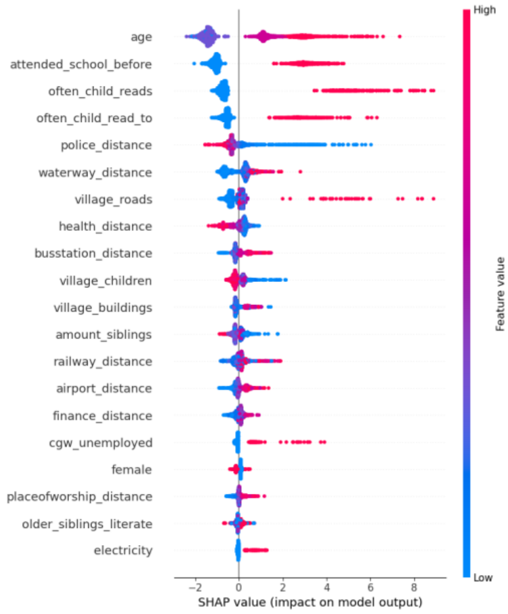
(EGRA baseline scores)

Outcome: Numeracy

(EGMA baseline scores)

Child + Village combined (Exp 3 & 6)





RQ2: Which factors best predict improvements in foundational learning outcomes after implementing an educational technology intervention with out-of-school children in remote areas of Tanzania?

Model performance

Table 13 reports the results of the six experiments conducted to identify factors that best predict improvements in foundational learning, at the end of the field trial, after implementation of the educational technology interventions. Once again, R2 scores were low (between 0.09 and 0.28) and the MAE were also higher than baseline predictor models (see Table 12), indicating that the model is relatively poor at predicting learning improvements. However, as Table 13 shows, model performances were an improvement on the dummy model, illustrating they have enhanced predictive capabilities. Performance scores indicate that once the app intervention has been implemented, village-level factors become more important and predict children's learning outcomes relatively more accurately than child-level factors. When child- and village-level factors are combined, model performance improves marginally above village factors alone.

Table 13

Summary of the performance of the six regressor models conducted to determine factors that best predict improvements in foundational learning skills after implementing an Edtech intervention with out-of-school children in remote areas of Tanzania (RQ2).

Dependent Variable	**Model Feature Set / Experiment**	**Dummy Performance**	**Best Performing Model Class**	**Model Performance**
Literacy (improvement)	Child (Exp 7)	R2: 0.00 MAE: 43.94 MSE: 3003.16	XGBoost	R2: 0.09 MAE: 39.93 MSE: 2752.13
	Village (Exp 8)	R2: 0.00 MAE: 44.09 MSE: 3050.57	XGBoost	R2: 0.20 MAE: 36.15 MSE: 2426.73
	Combined Child + Village (Exp 9)	R2: 0.00 MAE: 44.19 MSE: 3065.44	XGBoost	R2: 0.21 MAE: 36.02 MSE: 2444.99

Dependent Variable	Model Feature Set / Experiment	Dummy Performance	Best Performing Model Class	Model Performance
Numeracy (improvement)	Child (Exp 10)	R2: 0.00 MAE: 14.85 MSE: 346.56	Random Forest	R2: 0.14 MAE: 13.62 MSE: 299.32
	Village (Exp 11)	R2: 0.00 MAE: 14.82 MSE: 343.42	Random Forest	R2: 0.27 MAE: 12.25 MSE: 253.89
	Combined Child + Village (Exp 12)	R2: 0.00 MAE: 14.90 MSE: 345.24	Random Forest	R2: 0.28 MAE: 12.27 MSE: 251.92

Variable importance

Figure 10 reports SHAP summary plots for each of the six experiments conducted to determine the best predictors of improvements in foundational learning outcomes after implementation of an educational technology intervention in remote villages in Tanzania.

At the child level, the strongest predictor of improvements in foundational learning was participation in the control group, as children who participated in the control group were shown to make the least improvements in foundational learning skills over the course of the field trial compared to children who received one of the app-based interventions. However, which app children received had a relatively greater influence on the prediction of improvements in numeracy than literacy skills; not surprisingly the two winning apps were highly and positively associated with improvements in learning. Gender was a strong predictor of literacy but not numeracy learning outcomes, with a positive relationship demonstrating that girls made relatively greater gains in literacy than boys. Similar to baseline results, reading and being read to had a clear positive relationship with improvements in learning outcomes, although this was only marginal for numeracy. Number of siblings was also associated with improvements in foundational learning, as children that had fewer siblings made greater improvements in foundational learning than children with many siblings. Furthermore, results showed that having more younger siblings that were literate had a marginally negative impact on model performance, whereas having more older siblings that were literate positively impacted model performance.

At the village level, relationships were again less obvious than at the child level, and mainly showed similarities to those found at baseline. Villages in close proximity to a police station or health service were shown to have greater improvements in foundational learning than those further away. The strongest predictor for numeracy outcomes was distance to an airport, where villages close to an airport were associated with less improvement in foundational learning than those further away. Again, reachability had a low impact on model performance overall, but very hard to reach villages were marginally associated with poor improvements in numeracy outcomes.

Figure 10

SHAP summary plots for the six experiments conducted to determine factors that best predict improvements in foundational learning skills after implementing an Edtech intervention with out-of-school children in remote areas of Tanzania (RQ2).

Outcome: Literacy

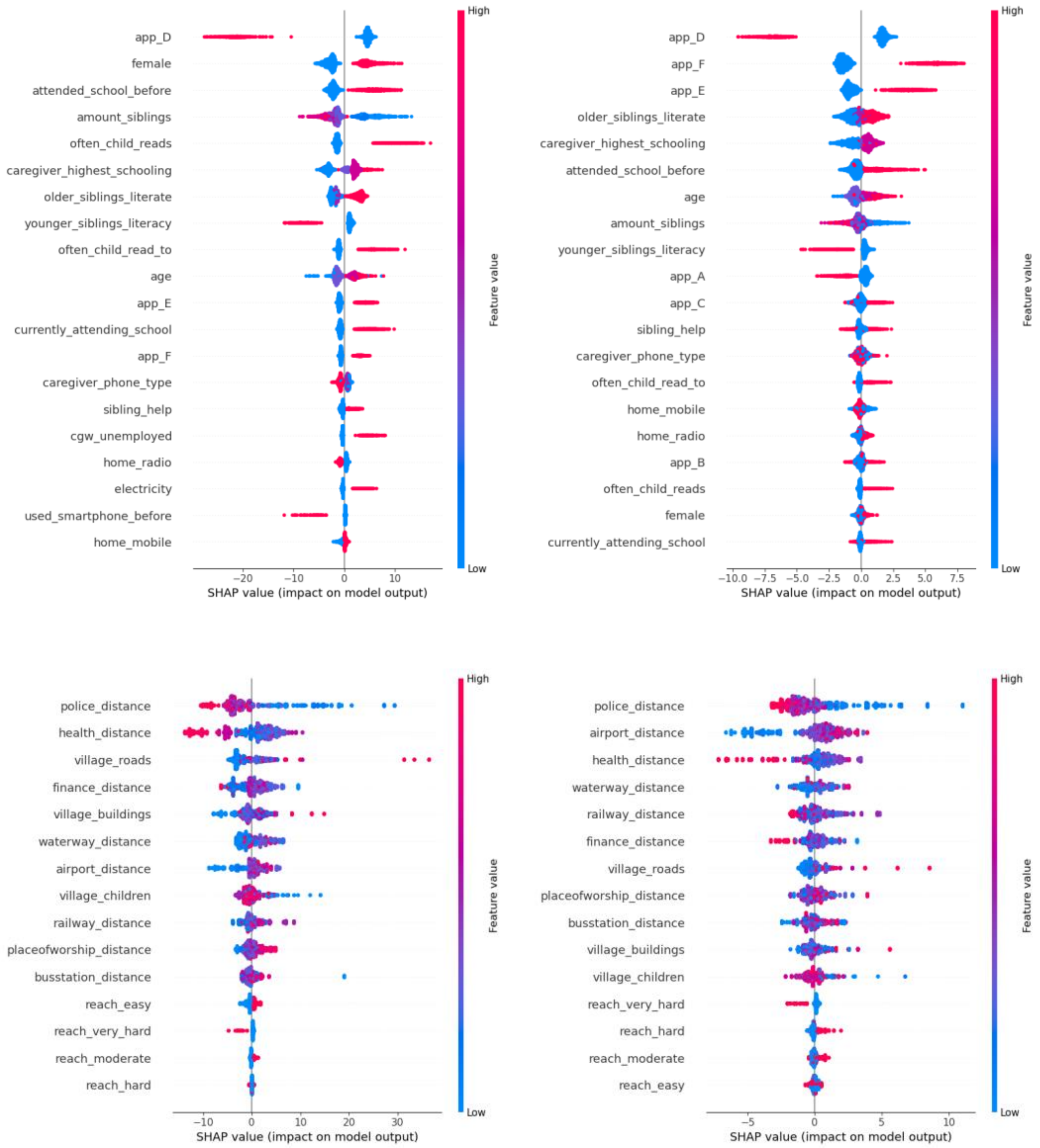
(EGRA improvement scores)

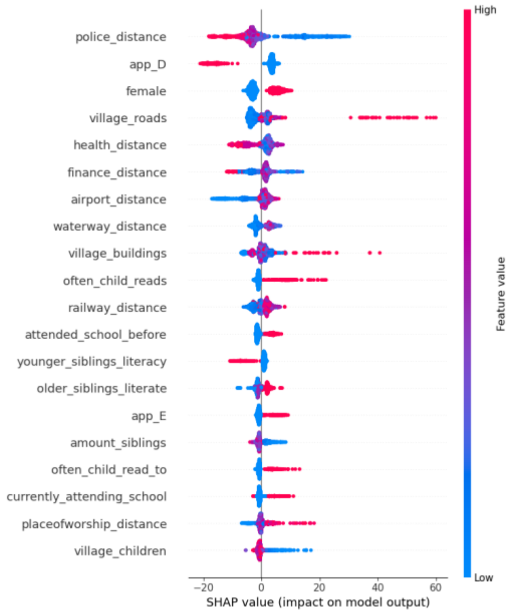
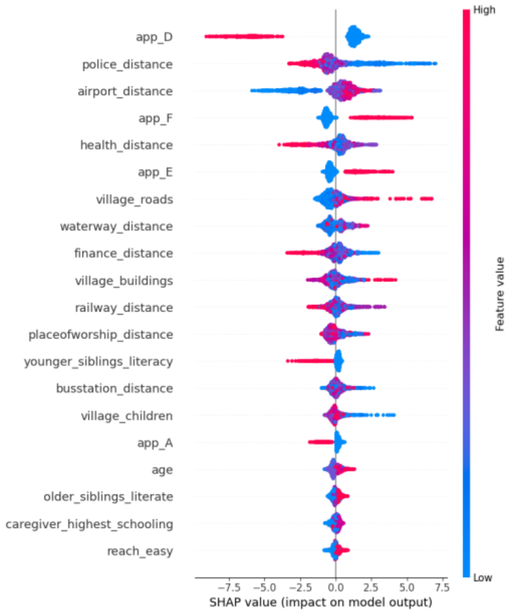
Outcome: Numeracy

(EGMA improvement scores)

Child (Exp 7 & 10)

Village (Exp 8 & 11)



	Outcome: Literacy (EGRA improvement scores)	**Outcome: Numeracy (EGMA improvement scores)**
Combined child and village (Exp 9 & 12)		

Discussion

This study reports twelve experiments designed to identify contextual factors that best predict foundational learning in out-of-school children in remote areas of Tanzania, before and after implementing an educational technology intervention. Rurality has been shown to be a significant predictor of educational outcomes in LMICs (e.g. Spaull, 2012; UNESCO, 2015), but until now, the mechanisms driving poor educational outcomes in remote contexts have yet to be elucidated. To address this gap in knowledge, for the first time, machine learning methods were compared to traditional linear regression to explore which type of model best captured data comprising child and village level features on foundational learning outcomes. Machine learning models were explored as they avoid common pitfalls with linear models when considering large and complex datasets, such as overfitting and multicollinearity (Lavelle-Hill et al., 2021). Results showed that for each of the experiments conducted, a machine learning model produced a better fit of the data than traditional linear regression.

Overall, a machine learning model produced a better fit of the data than the dummy model in all instances, demonstrating that the machine learning model produced the most accurate prediction of children's learning scores. Of the different machine learning models tested, the best-fitting model class was either Random Forest or XGBoost for each experiment. These results are important as they demonstrate that traditional linear regression is not suitable for large and complex datasets that are typical in different areas of developmental research, including education. Future research for development that analyses large and complex datasets need to adopt machine learning methods, as utilised here, to capture data patterns accurately.

The machine learning methods employed here also enabled variable importance to be determined. Results showed at baseline, prior to intervention with an educational technology intervention, child-level factors best predicted learning of foundational skills. In contrast, after an educational technology intervention had been implemented within the villages that took part in the GLXP competition, village-level factors that represent physical isolation and social connectedness were shown to be most predictive of improvements in learning of foundational skills. This novel finding is important as it highlights for the first time how geospatial and environmental features contribute towards the success of educational technology interventions that are implemented directly to the community in low resource settings.

This has important theoretical and practical implications. Initial results from the child-level investigations (experiments 1, 4, 7 & 10) corroborate previous research and thus serve as validation for the application of machine learning models in the analysis of data from the GLXP competition. Across the four child-level experiments, prior school attendance was shown to be one of the most influential predictors of successful learning outcomes for out-of-school children, both before and after a technology-based learning intervention had been implemented. This is consistent with previous research indicating that prior knowledge significantly impacts knowledge acquisition in LMICs, with and without technology use (Tavera & Casinillo, 2020; Gulliford & Miller, 2023). Furthermore, reading, and being read to, were significant predictors of learning, particularly at baseline for literacy and numeracy. Reading remained a strong predictor of improvements in literacy skills, but became less important for learning numeracy, after an app-based intervention had been introduced. This finding corroborates previous research that has highlighted the importance of the home reading environment and shared reading on academic success (Timmons & Pelletier, 2015; Jeynes, 2012).

Some findings were, however, unexpected and warrant careful interpretation. For example, at baseline, having an unemployed caregiver was a higher predictor than any other profession, particularly impacting numeracy outcomes. This result is somewhat surprising as the caregiver profession was used as a proxy for parental SES and previous research has shown that high parental SES is typically associated with high academic achievement (Long & Pang, 2016; Zhang et al., 2020). It is, however, plausible that within the context of this study, unemployed caregivers might have had more time available than parents of other professions to dedicate to supporting their child's learning. Previous research has shown that motivational support from uneducated caregivers can be beneficial for out-of-school children when learning foundational skills (Huntington et al., 2023a; see also Chapter 6). It is also probable that even in remote villages in Tanzania, unemployed caregivers are more likely to possess some basic numeracy skills compared to basic literacy skills, as basic numeracy is needed for everyday transactions, such as buying or selling goods at the market, and out-of-school environments have been shown to primarily contribute to illiterate numeracy development (Putrawangsa & Hasanah, 2018; Vandermaas-Peeler, 2008; Njie, 2016). This might provide an explanation as to why having an unemployed caregiver impacted numeracy outcomes at baseline more so than literacy outcomes.

In addition, few siblings was a consistent predictor of learning outcomes and this further consolidates the notion that time available by caregivers to support their child's learning may be influential in determining learning outcomes, either due to the employment status of the caregiver or the number of children in the household. Whilst this hypothesis requires testing empirically, previous research has suggested that being an only child can have academic advantages over children with siblings (Wei et al., 2016; Zhao et al., 2022), and that literacy and numeracy acquisition is negatively associated with the number of siblings a child has (Gurgand et al., 2023). Moreover, results from the village-level investigations also corroborate this hypothesis, as in all four experiments, low numbers of village children was associated with high learning outcomes, and number of children in the village was a significant predictor for numeracy at baseline. At first glance, this result might seem unexpected, as number of children in a village was a representation of potential peer networks, and research has shown that peer interactions can be crucial for children's cognitive skills and early development, particularly for out-of-school children lacking formal instruction (Garris et al., 2018; Tauson & Stannard, 2018). However, this finding could be indicative of the opportunity to learn from adults; having few peers and/or siblings could be beneficial as it provides a good opportunity for learning support from adults in the household and broader community.

Based on the available literature, this appears to be the first study to examine how village-level geospatial features are associated with foundational learning outcomes, either before or after intervention with an educational technology intervention. Across all experiments that explored village-level factors, proximity to a police station or health centre were shown to be key predictors of learning outcomes. These findings were anticipated, as distance to local services represented social connectedness, and feelings of social isolation are associated with poor educational outcomes (Leigh-Hunt et al., 2017). Physical isolation also impacted foundational learning outcomes. Results showed that before and after intervention with an educational technology, the number of roads in a village was positively related to learning outcomes, especially for literacy. This is consistent with previous research that has shown accessibility to be a significant barrier to participation in formal and non-formal education in Africa (Ogbonna, 2015; Ayiro & Sang, 2016). However, rurality and road quality, as measured by reachability judgments by the XPRIZE team, was found to be the least predictive village-level factor of learning outcomes, with little to no impact on the model fit. This is somewhat surprising as previous research with this sample has indicated that having accessible routes to charging points was crucial for the successful implementation of the educational technologies trialled in the GLXP competition (Huntington et al., 2023a; see also Chapter 6). Other research has suggested that being close to transport links may alleviate the negative impact of transport poverty on access to key services, such as education and health (e.g. Lucas, 2011; Porter, 2014). However, in the current study, most factors representative of transport links had minimal impact on learning outcomes, particularly for literacy. Indeed, living close to an airport negatively influenced learning improvements, especially for numeracy. This may reflect high levels of noise disturbance from living close to an airport or a lack of educational support from caregivers or community members if they worked at a nearby airport. Clearly, further research is needed to examine why these geospatial relationships with foundational learning outcomes exist.

Results also showed that the type of instruction given over the course of the GLXP competition was influential in predicting improvements in foundational learning. Children assigned to the control condition, who did not receive intervention with an educational technology, made little or no improvement in foundational learning over the duration of the field trial. In contrast, the app intervention that children received significantly impacted learning improvements, particularly if the child was assigned one of the two apps that were awarded joint winners of the GLXP competition. This validates the competition results, and further indicates that the combination of six key pedagogical features embodied in the two winning apps facilitates foundational learning for out-of-school children (see Chapter 3).

Whilst gender was not a strong predictor of baseline learning outcomes, it was a strong predictor for improving foundational skills, as girls were shown to make greater improvements than boys in literacy and numeracy after receiving 15-months intervention with an educational app (although the association was marginal for numeracy). This is consistent with the findings in Chapter 4 with this demographic. It also supports previous research that has argued that girls can benefit from learning if they have access to education, with higher reading levels than boys and EdTech interventions resulting in girls catching up to boys in numeracy (Miheretu, 2019; OECD, 2016; Pitchford et al., 2019). The current results demonstrate that with prolonged implementation, girls can achieve greater improvements than boys in foundational learning, especially literacy, when receiving instruction with high-quality educational apps. This has important implications for addressing the gender gap in education that is found globally (Webb et al., 2020; EdTech Hub, 2023) and demonstrates how educational technologies can enhance opportunities for girls to learn core foundational skills.

As anticipated, familial literacy, as measured by the highest level of schooling of a caregiver, and how many of the child's siblings were literate, was found to be a strong predictor for improvements in foundational learning. Results revealed a clear and positive relationship illustrating that the greatest improvements in literacy and numeracy were associated with caregivers who had high levels of schooling, perhaps reflecting more parental involvement in their child's education than caregivers with lower levels of schooling or heritability of reading abilities (Gimenez-Nadal & Molina, 2012; Van Bergen et al., 2017). This result contrasts, to some extent, with the findings at baseline concerning the professional status of caregivers, as it might be assumed that caregivers with high levels of schooling are more likely to be employed than caregivers with low schooling levels. Clearly, further research is needed to unpack these findings, but it is possible that more educated caregivers were able to scaffold their child's learning through the support of an educational app to a greater extent than caregivers with low levels of schooling.

Similarly, a complex pattern of results was revealed in relation to the age of siblings and sibling's literacy level and the learning outcomes of children that participated in the GLXP. Results showed that households where older siblings have literacy skills positively influenced model performance and was a moderately important predictor of learning improvements, perhaps indicating that the older siblings helped support the children's app-based learning. This supports previous research in low-income countries that argued that high literacy of older siblings has a positive impact on a child's literacy development, as there is the opportunity for shared literacy experiences (Lindskog, 2011; Sokal & Piotrowski, 2011) and the potential for educational spillover effects (McCarthy & Pearlman, 2022). Moreover, Sokal & Piotrowski (2011) found that children from large families (three or more children) were less likely to experience shared reading with their siblings than children from small families, further demonstrating the benefits of having few siblings or few peers in the village, as shown at baseline in the current study.

In contrast, results showed that having younger siblings with literacy skills negatively influenced child improvements in foundational learning. This finding is inconsistent with previous research that has argued that the presence of siblings of all ages can promote literacy, working memory, and cognitive development (Workman, 2017; Knoester & Plikuhn, 2016). Siblings close in age have also been thought to play a crucial role in a child's literacy development through scaffolding and collaboration, leading to unique reciprocity between both children as they complete learning tasks together (Beffel et al., 2021; Knoester & Plikuhn, 2016). However, the GLXP implemented personalised learning technologies that were designed specifically to promote autonomous learning through individual child interactions directly with the software. Hence, there would have been little scope for collaborative learning with the software, so the introduction of these educational technologies for out-of-school children might have mitigated potential benefits of collaborative learning with siblings and peers. In addition, it is possible that parents encouraged younger siblings to engage with the tablet interventions upon realising its potential, which could have directed time interacting with the tablet away from the intended child. These suppositions require verification through future research.

Features related to prior experiences with technology (e.g. child has used a tablet before, smartphone before, and caregiver has used a similar device before) were the least important predictors of learning outcomes across all experiments. This is somewhat unexpected, as it could be assumed that previous exposure to technology may help familiarise a child and boost their confidence in handling a tablet. However, a floor effect might underpin this result, as only a few children and caregivers in this study had previous experience of using a tablet or smartphone. Likewise, help from an adult in the household or a 'village Mama' were also shown to have little impact on a child's improvement in foundational skills, after intervention with an educational app. This was surprising, as previous research from the GLXP identified support from village Mamas and other community members was plentiful and beneficial to the successful implementation of the educational technologies within the participating villages (Huntington et al., 2023b; see also Chapter 6). It also refutes broader research demonstrating the value of familial and community support in positively influencing children's learning outcomes in low-income countries (Tauson & Stannard, 2018; Byerengo & Onyango, 2021). These apparently conflicting results may reflect the nature of the educational intervention implemented in the GLXP that were designed to promote autonomous learning by individual children interacting directly with the software.

Limitations

A potential limitation of this research is the low performance of the best-fitting models in all experiments conducted, indicating that the overall model fit did not explain much of the variance in children's learning scores. Large MAE values were also found, particularly when predicting improvements in learning, which limit the extent to which inferences can be drawn. However, low R2 scores are commonly observed in educational research (e.g. Allen et al., 2014), which raises the possibility that there are other contributing factors to foundational learning that were not investigated here, as suggested in Chapter 4. For example, machine learning research by Smith et al., (in prep) has shown dietary diversity and soil type to be strong indicators of primary school education outcomes in Malawi and malnutrition can lead to cognitive deficits and impaired brain development in early childhood (French et al., 2020). Reliance on pre-existing data precluded exploration of nutritional factors on learning outcomes in the current study, as existing data on nutrition in Tanzania was only available at the ward level, rather than the more granular village level. However, the role of nutrition in predicting foundational learning outcomes, with and without EdTech interventions, with out-of-school children in remote settings is clearly an avenue worthy of future research.

Additionally, there are limitations when using community-sourced geospatial data from resources such as Open Street Maps, as data accuracy and quality cannot be determined, and urban areas are updated more regularly than rural areas (Vargas-Munoz et al., 2020).

However, there has been a growing focus on Tanzania in geospatial research over the past decade, as demonstrated by the Building Footprints project as part of the AI for Humanitarian Action programme, in which Bing Maps and Microsoft Philanthropies released 11 million country wide open building footprints datasets (Microsoft Bing, 2020). Any of the ten million users on OpenStreetMaps can correct obvious errors and the geospatial data utilised in this study was also last updated shortly after completion of the GLXP competition, so it was as accurate as possible, considering it is based on community contributions. Using community-generated data is common in geospatial research as it is the most practical alternative to manually collecting data at the ground level.

Implications and conclusions

This study has important implications for theories of how educational technologies can support learning of foundational skills, which should inform the design and implementation of effective educational technologies to improve educational outcomes for out-of-school children in low-resource settings. This study has highlighted the importance of contextual factors at the child and village level that are associated with the learning of foundational skills by out-of-school children in Tanzania with educational technologies. This adds to the knowledge base from previous intervention research with educational technologies which has focused predominantly on individual differences (e.g. Bardack et al., 2023; Outhwaite et al., 2020; Lurvink & Pitchford, 2023), or contextual factors within a classroom setting (e.g. Gulliford et al., 2021; Outhwaite et al., 2019b).

For out-of-school children, this study has clearly shown that factors relating to the child’s home, social, and physical environments are primary indicators of successful learning outcomes, and highlight the need for technology designers, governments, and implementing agencies to consider contextual factors when developing app-based learning interventions within LMICs, if they are to be used effectively to address the global learning crisis.

Chapter 6: Expert Elicitation of the GLXP competition

Background

As highlighted in Chapter 1, children in LMICs who are receiving no education or low-quality education might benefit from a learning app intervention to acquire core foundational skills. However, the effectiveness of learning technology remains under-researched and under-documented in out-of-school children, particularly in Sub-Saharan Africa. As a result, it is unclear how EdTech might be deployed and implemented successfully directly to communities to promote the learning of foundational skills in out-of-school children. Chapter 5 investigated the child-level and community-level predictors of foundational learning skills to identify which contextual factors are most important when considering an EdTech intervention with out-of-school children. It was also established that contextual factors were more influential than child-level factors in promoting foundational learning skills, emphasising the importance of infrastructure and good-quality implementation practices. This chapter builds on those findings further by using the experiences of the GLXP competition to explore the potential of technology to support autonomous learning among out-of-school children in remote villages - a population who are deemed among those most at risk of experiencing the profound effects of learning poverty (Pitchford & Outhwaite, 2016b; UNESCO, 2022c). Expert perspectives from the GLXP were sought using semi-structured interviews to generate insights into the potential of EdTech interventions to mitigate the effects of the learning crisis in Sub-Saharan Africa, particularly in remote villages. Furthermore, the challenges and opportunities associated with implementing these interventions were explored, providing valuable information for stakeholders involved in educational policy and practice in these contexts.

Autonomous learning for out-of-school children through EdTech

Educational apps promoting learner autonomy are arguably a pragmatic solution for facilitating learning for out-of-school children, where there is no access to traditional schooling, as the learning style is child-centred, and an experienced adult is not required to scaffold the learning process. Autonomous learning is already a central feature of many educational apps, enabling the child to take control of their learning, supporting the management of what they learn and when, and engaging reflectively in the learning process (Lan, 2018; Chapter 3). Cultivating and encouraging learner autonomy may be critical to improving children's intrinsic motivation, sense of hope and agency, and educational outcomes and providing a different type of scaffolding for children with different or additional needs (Outhwaite et al., 2019a; World Bank, 2021). As demonstrated in Chapter 1, promoting autonomous learning using EdTech has garnered successful outcomes in low-income countries during the COVID-19 pandemic (World Bank, 2021) and with children that do not typically attend school or have never accessed formal education before (e.g. Stubbé et al., 2016; Gottwald et al., 2017). The autonomous learning process has improved foundational learning gains and enhanced motivation to learn, demonstrating that EdTech has the potential to be effective for out-of-school children (Gottwald et al., 2017).

This emerging evidence highlights the potential for educational apps to address the global learning crisis. Whilst some studies have shown positive gains start to diminish over time (Tauson & Stannard, 2018), assessments with a small sample of children in Ethiopia showed encouraging results for reading comprehension, word decoding and reading texts after one year (Gottwald et al., 2017). Clearly, further research is needed to identify and address factors influencing the sustainability of effective EdTech interventions with out-of-school children (Walton, 2018). It has been argued that while out-of-school children can learn autonomously, community members are crucial in facilitating and guiding children through the learning process, even if they are low-skilled, and predominantly for purposes of encouragement (Stubbé et al., 2016; World Bank, 2021). Children in these contexts may lack the necessary experience and cognitive skills to achieve productive independent enquiry, so would strongly benefit from teaching and assistance in learning independently before being left to do so (Dean & Kuhn, 2007; Paradowski, 2015). For EdTech interventions to succeed with out-of-school children living in remote, low-income settings, implementers first need to establish if autonomous learning is possible; and if not, what solutions can be found to help scaffold children's learning with the resources available.

Current study

This study aims to advance understanding of how autonomous learning can be achieved and sustained with out-of-school children from low-income remote settings when given a tablet equipped with an educational app. The quantitative results of the GLXP imply that educational apps deployed directly within remote villages can support autonomous learning (XPRIZE, 2019). However, qualitative data regarding implementation was not gathered during the competition, so without knowing what happened on the ground, it is not possible to draw firm conclusions about interactive apps promoting autonomous learning. Further insights are needed from key informants of the competition to determine if additional factors might have influenced the successful implementation of the learning apps within the participating villages and, ultimately, the learning gains achieved. Post-intervention research can reliably inform long-term implementation and sustainability and provide insights into the general acceptability of an impact evaluation within the target audience (Pegrum et al., 2013), yet very few qualitative studies have been reported in this context for that purpose.

To investigate factors contributing to the implementation process of educational apps deployed directly within remote villages for out-of-school children, an expert elicitation was conducted using semi-structured interviews with key informants of the GLXP. Expert elicitations are a “structured approach for obtaining judgements from experts” (p.133), usually conducted about items or events of interest to inform best practices for policymakers (Verdolini et al., 2018). In this study, data generated from the semi-structured interviews were subjected to Thematic Analysis due to its highly flexible, impartial nature (Clarke & Braun, 2014). The experiences and perspectives of the 14 experts associated with the GLXP were carefully analysed to generate core common themes. Accordingly, this exploratory chapter investigated the following research questions:

1. How do experts associated with the GLXP perceive the impact of EdTech in supporting autonomous learning in remote low-income settings?
2. What were the key challenges and opportunities identified by experts during the implementation of the GLXP competition?

Method

Research Design

Semi-structured interviews were conducted to gather qualitative data on the experience and perspectives of 14 experts associated with the GLXP. Semi-structured interviews with predetermined areas of focus afforded a valuable guiding framework for information elicitation without restricting the interview scope (Howitt, 2016). This interviewing technique allowed flexibility for both the researcher who may wish to probe further on points of interest and participants who may wish to discuss additional topics that they deem important to the research (Kallio et al., 2016). This adaptability also allowed for an informal, conversational style, enabling participants to discuss and share their experiences comfortably and openly (Bryman, 2016).

The qualitative analysis presented in this chapter (including data collection) was conducted before the statistical analyses detailed in Chapters 3-5. The qualitative data remained independent, and the analysis was thus not influenced by any findings from the quantitative studies, allowing for an unbiased, inductive interpretation of the data in this chapter.

Ethics approval

Ethical approval was granted by the School of Psychology Ethics Committee at the University of Nottingham (ethics reference: s1247). Informed consent was obtained from all interviewees in line with the British Psychological Society guidelines.

Participant Recruitment

Interviewees were recruited using purposive and snowball sampling strategies (Robson, 2002). Initial interviewees were selected in collaboration with the Executive Director of the XPRIZE Foundation. Further interviewees were identified through snowball sampling when conducting interviews. The researcher asked interviewees if there were any further individuals that they considered to be crucial to the competition. Recruitment ceased when all avenues had been exhausted, and the interviewees' suggestions were of previously interviewed individuals.

Participants

Fourteen individuals with key roles in the GLXP participated in the study. Interviewees had a mean age of 46.77 years (SD = 10.45, range 33-68); there were nine males and four females. Eleven participants lived in the United States, one in Tanzania, and one in the Netherlands. One participant refused to disclose their demographics due to concerns about anonymity. Participants had different job roles within the GLXP, including senior members of staff from the XPRIZE Foundation, members of the finalist app teams, data analysts, judges of the competition, and individuals who directly oversaw the technological implementation on the ground. At least four participants spent direct time in the participating communities during the competition. Any other captured demographics are not reported to protect anonymity in a specific participant pool. Participation was voluntary; no incentives were offered for taking part.

Data Collection Procedure

Recruitment emails outlined the study aims and potential time commitments and provided a detailed information sheet. All interviewees were given access to the semi-structured interview protocol before the interview. Where possible, a rapport was established between the researcher and interviewee via email before the interview took place.

Data collection took place over seven consecutive months (May - December 2020) via online video platforms Zoom (n=13) and Skype (n=1). The average interview duration was 81 minutes, and each interview was audio recorded by the researcher following informed consent from the interviewee. To address the two research questions posed by this study, the semi-structured interview focused on the perceived impact of implementing EdTech in low-income settings to address the global learning crisis (RQ1) and the challenges faced with the implementation process during the competition (RQ2). Additional information was gathered on the interviewees' job roles and other demographics (see Participants) and their future dissemination plans.

Data Transformation

Audio data from each interview was recorded, encrypted, and stored securely on the researcher's computer before transcription by hand. The transcription process is considered "a key phase of data analysis within interpretative qualitative methodology" (Bird, 2005, p227) due to early interpretations and meanings that might be conferred. Any personally identifying information was removed or de-identified, and participants were numbered for reference. A file containing the participants' identifying information was saved and stored separately.

Data analysis

Reflexive Thematic Analysis was conducted systematically using a six-stage process specified and revised by Braun and Clarke (2006; 2021). As this study was exploratory, an inductive approach was used in which no pre-assumptions or hypotheses were placed on the data based upon a particular theoretical stance. In line with the realist approach adopted in this analysis, themes were identified only at a semantic level (Braun & Clarke, 2013).

Transcripts were actively read and re-read numerous times by the researcher to become familiar with each account, and immersion in the data allowed initial analytic thoughts to develop. The second stage was to identify initial codes which applied across the data. Codes were used to label semantic or latent content within the data that was organised into meaningful groups (Maguire & Delahunt, 2017). These codes were then used to generate initial themes, which were reviewed and iterated upon throughout the analytical process. Developed themes were then refined, defined, and named appropriately in a manner that best captured the essence of what the theme conveyed before the final step of writing up the themes for dissemination.

All data analysis was conducted independently, in discussion with the supervisory team. Coding reliability was not necessary as in reflexive Thematic Analysis, meaning and knowledge are understood as contextual, and the subjectivity of the researcher is considered a resource to develop the production of knowledge rather than a threat to credibility (Braun & Clarke, 2019).

Researcher characteristics and reflexivity

The researcher is a doctoral student with a strong interest in educational development and prior university-based experience in pedagogical research and conducting qualitative research. The supervision team are academics in educational technology research and data science for social good; both have published widely about marginalised communities in Africa. To ensure reflexivity, the researcher kept a journal of field notes throughout the interviews documenting their initial reactions, thoughts, and judgements (Trainor & Bundon, 2020). This process also included the mood and context of the interview, how the researcher had influenced the interview, initial interpretations, and anything further that could be incorporated into the interview schedule. When coding a transcript, the corresponding entries were examined to provide further context and develop analytical understanding. While coding, further notes were made to record early interpretations and gradually develop coding decisions (Levitt, 2018).

Analysis & Preliminary Discussion

Four superordinate themes were generated from the experts' qualitative data that addressed how they perceived the impact of the EdTech deployed in the GLXP competition in supporting autonomous learning (RQ1), and key challenges and opportunities that implementing EdTech directly to communities might afford (RQ2). These were: 'Technology as a novel concept', 'Children don't learn in a vacuum', 'Respecting the cultural context' and 'Accessibility problems in a mobile world'.

Theme 1: Technology as a novel concept

The first theme had two subthemes, 'Wariness of the unknown' and 'Interest in new digital tools', and they encompass the concept that tablet technology was novel to many children living in remote villages in Tanzania. Relevant extracts are highlighted in Table 14.

Table 14

Quotes to support Theme 1: Technology as a novel concept

Subtheme	Quote	Number
Wariness of the unknown	"There were some concerns from the villages around them thinking this project was associated with freemasonry and you know what are these tablets you're trying to bring into our village" (P14)	Q1
	"Knowing the history of colonialism in the area makes total sense in terms of what are these people and knowing it's a foreign country" (P7)	Q2
	"They were just very suspicious. Like what is this tablet gonna do again. Will the tablet listen to what's happening in the house, yeah, where's the sound going to you know? Is this a thing that spies on us? […] You know if you introduce something new and this is like a super new area to them, yeah people tend to step back. You have to be a really into your innovation type of chief to allow this to be happening in your village when you don't know what it involves." (P2)	Q3
	"They may have never seen a touch screen before, and you know all of a sudden they have a piece of glass which is talking to them in their own language." (P13)	Q4
	"There were questions in certain communities about witchcraft you know was this witchcraft were we trying to do something there were some rumours about it would steal the soul of your	Q5

Subtheme	Quote	Number
	child. That one was widespread and like they've never seen it before I mean it is kind of witchcraft I don't know how a tablet works I'm like it seems magical that you can have something like that." (P7)	
	"Some of these children had not even seen a smartphone before" (P14)	Q6
	"They were using these tablets. And they heard the voices. They heard voices inside a tablet, so they broke open the tablets 'cause they want to know where the voice was coming from." (P11)	Q7
	"If a kid is curious and they're opening up the tablet to find out what's in there then that's your future engineer that kid gets another tablet […] we really were like this is just a curious kid who wants to know and they're opening it up and maybe they want to build it, so we encouraged that kind of curiosity." (P7)	Q8
	"those problems really melted away in the first few months, there were very little of [them] by a quarter way through our field test." (P7).	Q9
Interest in new digital tools	"There were some jealousy issues the biggest issue we heard back was this family have two tablets and we have none and that kind of thing." (P7)	Q10
	"For the children, it was excitement. It was not just, you know as school, but it was a new gadget. It was fun, it does games or whatever." (P3)	Q11

Subtheme	Quote	Number
	"They had things to share, they had things to tell, and their parents would say now I can talk with my child because they are so excited about the learning they have been doing and want to tell me about." (P3)	Q12
	"Some kids felt like they had already learned enough. They wanted something more." (P3)	Q13

Subtheme 1a: Wariness of the unknown

Interviewees emphasized that early in the trial, apprehension was felt and voiced by villagers approached to take part, and who did not initially know of the XPRIZE Foundation or the purpose of the field trial (Q1). The misconception of XPRIZE being a freemason organisation was rationalised as a potential link to colonialism (Q2). Interviewees felt villagers were initially sceptical about a Western organisation (XPRIZE) coming with an unfamiliar solution to their problems. There were also general suspicions about the intentions of the competition and those running it, and the capability of tablets to act as surveillance devices (Q3). While interviewees highlighted this concern from villagers, they also explained why this was an understandable challenge within the trial, due to the innovations involved in modern technology (Q4). They described how the villagers justified the tablets' abilities with the explanation of witchcraft and tried to relate to this with their own amazement (Q5). Interviewees empathised with the villagers' difficulty acclimating to such a novel concept. Despite modern technology, such as tablets, being prevalent in Westernised cultures for approximately 40 years (since the widespread introduction of personal computers in the 1980s; Abbate, 1999), villagers felt wonder that such advanced options were possible and at their disposal for teaching and learning.

Interviewees also reported children in the competition had difficulty with the concept of tablet technology and how it functioned (Q6), which resulted in tablets being broken (Q7). Whilst the breakages seem like a negative event, interviewees insisted that this was not the case (Q8). Breakage was explained as natural curiosity in children to be encouraged and celebrated, as children thrive on such curiosity during learning (Flannagan & Rockenbaugh, 2010). These breakages represent a desire to learn and an inquisitive nature, and interviewees felt nurturing this was important as it was a great indicator of potential success for children benefitting from the intervention. Whilst there was clearly a prevalent theme of wariness at the intervention, presented in varying levels of mistrust and curiosity, interviewees stressed that these issues were not pervasive and were gradually overcome with time, patience, and perseverance (Q9), a situation not dissimilar to when technologies are integrated into classrooms in the US (Couse & Chen, 2010).

Subtheme 1b: Interest in new digital tools

Once children realised what was happening, there was a sense of excitement as they were empowered to learn with the technology. However, due to the selectiveness of the competition, with only children of a specific age being targeted to receive the tablets, some children felt left out (Q10). This jealousy indicates that the tablets may have been considered a desirable amenity to have within the family, either because they are a source of entertainment; or because they reflect a high-value possession that might elevate social status within the community - particularly if the community are coming together to help support children that have them. Many interviewees indicated that children were extremely excited by this new method of learning (Q11). This suggests that the excitement was not just because children were learning (sometimes for the first time) but because the tasks, games, and technology were new. Children liked getting involved in something so different from their usual setting and learning whilst having fun. That interviewees emphasised how excited children were over the intervention indicates that uptake and interest in mobile learning would be high in remote settings. The intervention also enhanced children's communication skills with their families, even when parents were not personally involved in the learning tasks (Q12). For the first time perhaps, children had information to relay in which they were the 'experts'; they could tell stories about their experiences and bond with their parents by teaching them things they had learnt. However, one interviewee admitted that for some children, interest waned over the 15-month trial period (Q13). This suggests that once tablets become familiar technology to children and the initial novelty has worn off, learning may become less interesting and engaging for children. However, the interviewees indicated that children wanted more, suggesting this was driven by a desire to learn.

Theme 2: Children don't learn in a vacuum

The second theme contained two subthemes, 'It takes a village to raise a child' and 'Simulating the teacher role', and they encapsulate the idea that children do not, and cannot, learn in a 'vacuum' by themselves. All interviewees emphasised that despite the XPRIZE competition presenting the notion of 'fully independent' learning, this was simply not feasible or realistic (Q14). This theme is supported by the quotes shown in Table 15.

Table 15

Quotes to support Theme 2: Children don't learn in a vacuum

Subtheme	Quote	Number
-	"I've seen work that says they can just pick up a tablet and start using it tomorrow you know, and we know that's not true." (P1)	Q14
It takes a village to raise a child	"The village Mama became very important. They will talk to them in a way that is not parental, it is one of an adult, who is an educator, just not in the traditional sense." (P3)	Q15

Subtheme	Quote	Number
	"There was obviously a different level of the sort of support that the village Mamas provided, you could almost say some did their job and went home essentially and some sort of went the extra mile to support the children with some of the apps and stuff." (P2)	Q16
	"There was obviously a different level of the sort of support that the village Mamas provided, you could almost say some did their job and went home essentially and some sort of went the extra mile to support the children with some of the apps and stuff." (P2)	Q17
	"I've seen it with my own eyes like a lady opened up her shop and in the shop is nothing more than a table and a tent. And that is it and she would open it up for the kids so they could sit there or they would sit on the tree." (P2)	Q18
	"People would gather round erm but usually the kid that would have the tablet in most communities we saw even though we would maybe share it they wouldn't let other kids touch it they would be like I'll do the, you know, controls." (P7)	Q19
	"I did see kids setting up their own school, they would gather together under a big tree and meet at a certain point. It wasn't just siblings, it was a group of kids, and they were like 8 or 9 taking turns with the tablets and erm, yeah, there were several instances of that, and you know if they had a very active village Mama then she would you know, like encourage them too." (P2)	Q20

Subtheme	Quote	Number
Simulating the teacher role	"We rely on that idea that all children will love to explore freely too much. And we didn't know that some children actually had been trained to follow the instructions very strongly from early days. Our idea actually gave too much freedom for many learners who have never previously had a chance to explore themselves. […] If we did it again, we may make it more instructional." (P5)	Q21
	"It's certainly possible that the kids in our app went through the content just exploring it without really knowing what they were doing. Or maybe you know, maybe exhausted some of the content too quickly without fully digesting it." (P11)	Q22
	"There is a tendency of children to choose easy problems that they're going to get 100% correct and how do you navigate that in these informal learning environments where there isn't a teacher to push them harder and get them on that track." (P1)	Q23
	"What is needed is more support and I think motivation to engage in some of the other types of activities that they might find more challenging" (P8).	Q24
	"The [avatar] kind of mimics what the child's experience would be in a good school where they would get a lot of attention, so there's a teacher character. There's praise, there is guidance and structure, you know, there's repetition, there's practice, there's remedial work." (P13)	Q25

Subtheme	Quote	Number
	"That feedback and sense of progress are really important. […] Do that in a way that encourages persistence, and while building a sense of self-efficacy and giving people a sense of progress." (P6)	Q26

Subtheme 2a: "It takes a village to raise a child"

The XPRIZE competition focused on facilitating autonomous learning wherever possible. However, a common theme within the interviewees' narrative was that there were many indications that considerable communal support was offered to children during the learning process. Despite the Mama or Baba in each village being told not to assist in learning, they became crucial support to children in the field trial (Q15). Different levels of support were witnessed (Q16), with most interviewees emphasising that the Mamas were generally very hands-on, and provided support, guidance, and motivation to children when they were unsure on what they were doing or just wanted someone to show their progress to. While not as focused as the support children might receive from a traditional teacher, the Mamas' support was reported as instrumental in instilling confidence in children and providing them with a safe non-parental space to bring any problems to. Interviewees emphasised that educational background was not important (Q17).

Other adults within the wider community were also keen to help children in any way they were capable, indicating the presence of social interaction and caregiver support, which can be important for children's development (Q18; Hirsh-Pasek et al, 2015). Interviewees explained that communities were willing to support children to have a safe working environment in which to learn with the tablets, whether this was by helping children create a specific place to work or by helping them to travel to charging stations. Interviewees emphasising that they had seen these forms of support with their “own eyes” promoted the importance and pride they placed in such activities – and their expression of that support as an important personal account, that needs to be communicated and believed, was capturing the essence of what the community experienced during the competition.

Guidance from adults within the community was not the only support that interviewees reported. Children gathered in groups to play on the tablet together, suggesting that peer learning was important to them even when not enforced (Q19). The dominance over being in control of the device demonstrates feelings of possession and attachment to the tablet, with children considering the tablet theirs to share as they saw fit - rather than a commonly shared unit. Despite this, groups of children developed informal schools to facilitate shared learning (Q20). The tendency for a lot of children throughout the competition to “create a tablet school” (P7) indicates that despite having no formal learning context, children enjoyed a broader learning environment and orchestrated themselves to facilitate sharing and group learning. Seemingly, buildings and physical provisions are less important than the atmosphere and quality of group dynamics that supports and guides children to develop their abilities. This also allows the “child to teach [other children], which helps a child learn even better” (P3).

Subtheme 2b: Simulating the teacher role

The XPRIZE competition did not dictate the design or content of the software to participating teams, nor how to best encourage children to engage. This was a subject of much discussion, with interviewees emphasising the importance of balancing instruction and offering free choice “for feelings of autonomy” (P1). Some interviewees stressed that their teams initially felt they had miscalculated the amount of freedom that should be given to children - and perhaps had even conferred too much choice throughout the learning process (Q21). Interviewees suggested that the local culture instilled a sense of obedience in children, and while this was something that children were familiar with and comfortable with, it constrained exploration and self-directed learning. Some interviewees expressed the desire to improve upon this in the future, and others felt the high level of freedom negatively impacted the software solution's longevity (Q22). Interviewees also talked about the freedom provided by giving children more of a ‘browsing opportunity’ and worried that this may have distracted them from fully investing in the content as they would have hoped, as no one was telling them they had to do it. Interviewees considered the reasons that adding structure could benefit learners in out-of-school environments (Q23).

The idea of the teacher as the ‘motivator’ of children, and this element being absent in the software design, highlights one of the problems that participants felt was potentially prevalent within EdTech learning interventions in general (Q24). Although the interventions analysed in this study were heavily focused on autonomous learning, this was often at odds with what interviewees believed to be key to good practice in education. However, interviewees were aware of the context the competition was situated in, which lacked the option of teachers or formal schooling, with some teams actively attempting to address this by adding a teacher figure to their software – aiming to further motivate children and guide them appropriately through the content (Q25). Teams that included an avatar figure stressed the importance of giving children structured guidance that would benefit them most, aiming to help maintain learners’ interest while they progressed through the software at a pace suitable for themselves and one reflective of a classroom environment (Q26). Interviewees reported that to really fill the gap produced by the lack of a classroom, personalised and task-specific feedback needed to be given to encourage children not to give up, particularly with challenging tasks.

Theme 3: Respecting the cultural context

The third theme focused on the importance of respecting the cultural context to maximise success and avoid causing offence or risking failure based on a lack of contextual understanding. Table 16 shows supporting evidence for this concept.

Table 16

Quotes to support Theme 3: Respecting the cultural context

Quote	Number
"We understood the home environment where children are exposed to a lot of language, but there's no culture of reading so we understand where the child was at. We could have gone with a whizzy design with lots of sounds and colours and so on, and that has its place, but not in a context where children just aren't used to that sort of stuff. It needs to be as simple as possible and focus on the needs of the marginalised child."(P13)	Q27
"They often don't have the right cultural references to make them relevant to the particular learners, and that is an artefact of bringing in a bunch of content and games and things that have been successful elsewhere, so that's great, but erm you know successful for whom and in what context? So uh you know there's this race car game for kids who have very rarely even seen a car." (P1)	Q28
"We got some feedback as to like you know, why is this character doing this? Or you know this character doesn't seem to be a pleasant one or a fun one, which we thought they were. I think in one game we had an owl or something like that and then the feedback we got was that you know that is probably not appropriate as the owl has more negative connotations in Africa." (P9)	Q29

All interviewees emphasised the importance of being "confident of what opportunities there are for these learners in their environment" (P1) before creating the software and personalising it to the community of focus (Q27). The five finalists had various levels of previous experience with marginalised children, and children in Africa, so began the competition with differing levels of understanding and knowledge of the context they were working in. Some teams decided to keep things simple to avoid overwhelming the child with a novel environment and focus on making children aware of exactly what was happening within app flows. This highlighted a clear goal for some; understanding that children were already facing a novel situation simply engaging with the tablet hardware, there was increased motivation to keep other elements of the learning experience as basic and straightforward as possible. Other teams felt more confident about gamifying elements of the software and making them a challenge; yet others focused on producing tasks with rewards, hoping to keep children as invested as possible given their limited previous exposure to gaming technology. However, this came with its own identified challenges in transferring pre-existing game mechanisms to this new context (Q28). While for cost and time purposes, it can be helpful to exploit games that already exist (and have been shown to be successful in other contexts) interviewees stressed that using unfamiliar objects was not conducive to learning outcomes. One interviewee reported a hurdle their team faced was based on an character used in the software and the impact it had on children in remote villages in Tanzania (Q29), again demonstrating the importance of understanding the cultural context in which educational technology is to be deployed.

Theme 4: Accessibility problems in a mobile world

The fourth theme concerned accessibility problems during the competition that had an impact on children's ability to learn, as evidenced in Table 17.

Table 17

Quotes to support Theme 4: Accessibility problems in a mobile world

Quote	Number
"Some kids were small, and the charging station will be like over 3km and the kid can't walk with this gadget by herself to go charge it." (P3)	Q30
"There was one village where the kids had to go through an elephant area to get to the station. […] People were killed by the elephants, so those kids didn't dare to go to their charging station because they were afraid of the elephants." (P2)	Q31
"I love the thought of having a district bus decorated as a school and where kids can go with their tablets twice a week so they don't have to walk 10 miles, but the bus comes to them – that would be something and a possible solution." (P2)	Q32
"We made it so the tablets couldn't be used past 10pm. Households probably didn't have electricity, or another source of light so will play an app that just creates light so people can do other tasks, running down the battery and potentially not achieving much in terms of children learning." (P13)	Q33
"The tablets could break because they got water splashes on them or dust, or they were shaken or used too roughly" (P10).	Q34
"When it has run into problems, then you can't take advantage of industry to repair it because no-one's seen one before, there's no screen replacement that can happen on a local basis and so on and so forth um whereas you know, there are mobile phone dealers and technicians that are er you know more spread out around the country." (P1)	Q35

Whilst many infrastructural challenges were foreseen by the XPRIZE team and addressed wherever possible, some difficulties remained, for example, the geographical locations of the charging points for the tablets. Some villages were extremely remote so charging points were not within a convenient proximity (Q30). Charging stations were planned to be as accessible as possible, but due to the remoteness of some villages this was not always possible so imposed a significant physical and logistical strain on children and their families. Children's families also needed to see the value in learning with the technology, as families were burdened with organising visits to the charging stations when they were too hard to be accessed by the children alone. Interviewees, however, emphasised that villagers still made the effort to charge the tablets even when charging stations were far away, demonstrating a high level of interest in supporting learning with the technology.

There were not only physical limitations for children but sometimes additional safety challenges (Q31). This emphasises the importance of carefully planning infrastructure and demonstrates the benefits of detailed local knowledge, with different villages having different environments, routines, and surroundings that could impact on the success of using tablets in the community. Interviewees felt infrastructure was a problem that would be extremely difficult to overcome in rural locations, but some had started to think of potential solutions (Q32). However, the extreme lack of accessibility of some villages (with no clear paths or roads) may limit such solutions. Another interviewee explained that they knew it was impossible to change the competition's fundamental infrastructure, so they tried to adapt their software to fit the context (Q33). This would allow battery life to last longer and therefore reduce the frequency of trips to charging stations, limiting risk to children.

Another challenge raised was the fragility of the hardware itself, especially when not cared for effectively or kept in harsh conditions (Q34). Children were not familiar with looking after hardware of this nature, and despite being guided on how to do so, tablets often broke for practical reasons rather than out of curiosity as explored previously. This defines a further accessibility problem, not only due to hardware costs but also to long-term technology maintenance in areas that are unfamiliar with tablets (Q35). Even if funds were made available, interviewees stressed that knowledge and equipment was not currently available *in situ* to maintain a large-scale, ambitious undertaking of providing out-of-school children with tablet learning.

Discussion

This chapter explored expert perceptions of how EdTech might support autonomous learning with out-of-school children in remote settings in LMICs, and the challenges and opportunities that deploying EdTech within this context might afford. To achieve these aims, an expert elicitation was conducted with key informants from the GLXP competition to garner on-the-ground experiences to determine if the learning outcomes achieved throughout the competition could be attributed purely to autonomous app-based learning.

Four themes were identified that highlight a high level of community engagement in supporting the implementation and continuity of the EdTech intervention throughout the XPRIZE competition. Clearly, this challenges any inferences from the quantitative results from the GLXP competition that might indicate children can learn autonomously purely with interactive apps.

Technology as a novel concept

The first sub-theme showcased the associated confusion and problems caused by the novelty of the technology. Villagers were concerned about XPRIZE being a freemasonry, perhaps due to the documented chequered history of English and American colonialism in East Africa (Bulhan, 2015). This may explain why villagers were left feeling uncertain and wary about the involvement of XPRIZE if they were “sinister forces” that could be “plotting against the country” (The Economist, 2018, para.1). The further emphasis on spying suggests feelings of an insider/outsider divide and mistrust in newcomers introducing new technology that villagers had never experienced before. Belief in witchcraft has been perpetuated in many modern African cultures, particularly in circumstances where rational knowledge fails to explain an event or phenomenon, such as the existence of diseases with mysterious causes (Lewis, 2021). The novelty of something so new, different, and with no apparent rational explanation to the villagers may be why some initial attributions to witchcraft were made. This sub-theme highlights the importance of sensitization prior to introducing a technology intervention in low-income countries. The sensitization process should be gradual, thorough, and contextually bound for maximum success, with villagers being given the opportunity to explore any questions with experts.

The second subtheme demonstrated children's interest and excitement about learning with the tablet technology. This corroborates previous research on the use of EdTech for children in displaced settings, in which findings include increased engagement, motivation, and excitement (Islam & Grönlund, 2016). This study revealed valuable shared parent-child experiences that would likely consolidate learning for the child and their visible excitement, demonstrating the value of learning for their parents. EdTech has previously been shown to strengthen the bond and social interactions between parent and child, both during the COVID-19 pandemic and in multiple low-income emergency settings, as more opportunities to share and demonstrate new skills are experienced, allowing parents to further engage with their learning (Wang et al., 2020; Power et al., 2021).

Despite high levels of interest and excitement driving the use of tablets through the competition, it remains to be determined if this interest will be sustained long-term. Levels of interest in learning with the tablet technology decreased over the 15-month test period, and this has also been reflected in Western settings, where tablet use was initially preferred to 'traditional' learning methods but with interest waning following sustained use (Baytak et al., 2011; Muhammad et al., 2019). Motivation has also been shown to wane in children in developing countries over time as familiarity with technology develops and warns that the novelty of tablets is a temporary motivator that is not sustained (Gulati, 2008; Tamim et al., 2015). However, in remote environments where there are no other opportunities to learn, tablet use may continue long-term as this is the only viable option available to children. This possibility remains to be determined.

Children don't learn in a vacuum

The first sub-theme emphasised the extent to which families and the wider community supported children's learning experiences. Despite no formal instructions to get involved, the choice that some village Mamas and community members made to go the extra mile shows they valued children's learning and would work hard to support this. Emotional support and scaffolding are crucial for a child to progress and ensure that learning is productive, whilst reinforcing social interaction is needed for a positive learning experience (Hsin et al., 2014; Tauson & Stannard, 2018). Partnerships with family and community have previously been shown to positively influence children's learning in Tanzanian school settings (Byerengo & Onyango, 2021), so it is possible that establishing community-level partnerships could have a similar influence in out-of-school settings.

Children were also shown to receive support from peers. This is significant as peer learning can be crucial for children's early learning development, as it fosters collaborative learning, the development of meaningful social interaction, and enhances core cognitive skills (Garris et al., 2018). The creation of informal 'schools' or learning areas led some children to take on the teacher role, which is a common phenomenon when learners are given freedom and facilitates the teaching of new techniques to other children. This can, in turn, consolidate teaching-child learning and understanding. For out-of-school children lacking formal tutoring, such peer interaction could be highly beneficial (Tauson & Stannard, 2018).

Together, these results indicate that a fundamental contributor to the success of the Global Learning field trial was communities banding together to support children, practically and emotionally, despite many having never experienced formal education.

The second sub-theme showed that having an in-app avatar is important for out-of-school children who do not have access to a formal teacher (see also Huntington et al., 2023; Chapter 3). Some teams chose to provide children with lots of freedom, as previous research on autonomous learning has shown that children choose the most interesting, easiest, and most enjoyable tasks in the short term rather than tasks that will challenge them educationally (Couse & Chen, 2010). It has been argued that choice-based learning should remain integral, given the evidence children often learn best when selecting their own activity, direction, and pace (Chad-Friedman et al., 2019). However, Hirsh-Pasek et al. (2015) argued that an educational app should require thought, attention, and intellectual effort by the child, and interviewees agreed that the lack of structured guidance could be problematic for children's progress. Our study confirmed that in this context, the presence of an in-app avatar was a motivating influence and guide for children (Kolak et al., 2021). Major et al. (2021) emphasized the value of personalisation within EdTech to facilitate a learning experience driven by children's interests and needs, with support tailored as appropriate. The use of personalized feedback in EdTech significantly increases children's self-efficacy in learning, which is a key component for improving learning outcomes (Mouza & Cavalier, 2013; Outhwaite et al., 2023). Despite this, it has also been argued that scaffolding within educational apps is possible using feedback, learning in such remote environments may require conceptual feedback that engages children in far more complex, evaluative thinking (Cayton-Hodges et al., 2015). However, in a setting where literacy levels are extremely low for adults, as with the remote villages in the XPRIZE competition, this level of sophisticated scaffolding would most likely not be feasible from within the community.

Whilst autonomous learning was an ideological goal of the XPRIZE competition, interviewees noted challenges to free learning and emphasised the need to balance autonomy with directed progression through an educational app for it to be deployed successfully with out-of-school children.

Respecting the cultural context

The third theme centred around knowing and respecting the cultural context in which the intervention was to be implemented (Keengwee & Bhargava, 2013). A lack of developer knowledge of the cultural context was shown by some teams, illustrating practical challenges due to the negative impact of a game, character, or other elements within the app. Hirsh-Pasek et al. (2015) highlighted the need for app content to be culturally sensitive and the potential problems that arise when app content is not fully cognisant culturally. Children may be more likely to engage in culturally appropriate content than content not aligned with their context (Hsin et al., 2014). This will constrain scalability and challenges the notion that once shown to be successful, an app can be deployed as a global learning tool (XPRIZE, 2019). However, novelty *per se* might not be problematic, as children demonstrate an interest in games that include unfamiliar objects or scenarios (such as space explorations and imaginary play; Fleer, 2015; Suminar & Wardana, 2018). This suggests that app developers should consider the cultural context carefully when designing content and structure, ideally in consultation with communities that will receive the intervention, to ensure the longevity of use.

Accessibility problems in a mobile world

The fourth theme concerned accessibility with the mobile technology used in the remote villages of Tanzania, despite there being 4.32 billion mobile internet users globally (Statista, 2021b). Several infrastructural challenges were addressed prior to the competition, but issues commonly cited for mobile technology interventions in developing countries remained, primarily the distance to the charging stations and the safety of the children walking to them (Camfield et al., 2007; Crompton et al., 2021c). This supports the findings in Chapter 5, which demonstrated the importance of nearby amenities and services, suggesting that the location of the remote villages is pivotal for success. Previous technology interventions have attempted to circumvent accessibility issues by providing tablets with built-in solar panels. While this raises the cost of interventions, it lowers threats to safety and helps avoid logistical issues (Moss, 2020). Crompton et al. (2021c) considered digital access a fundamental barrier to equal education opportunities. This study revealed that the community was crucial in helping children access EdTech for learning.

Tablets were also shown to be utilised for other purposes, which is common in low-income contexts (Tauson & Stannard, 2018). For example, tablets were used for lighting in the evenings, which drained the battery and prevented children from accessing the learning apps. Another issue was hardware fragility, which is problematic in low-income, remote areas, as breakages incur high outlay and ongoing replacement costs over time. The cost of hardware is one of the greatest barriers to EdTech solutions to learning in low-income countries (Passey et al., 2016; Kaguo, 2011). Governments wanting to take advantage of digital technologies must consider how maintenance can be handled locally to maximise sustainability.

Limitations

A potential limitation of this study is the interviewees' investment in the success of the competition. There is the risk that some interviewees were more likely to maximise their successes and hide failures, as is common when reporting results in learning and teaching research (Dawson & Dawson, 2018). While interviewees' attitudes may be more positive than outsiders, results demonstrated that the interviewees identified several challenges. Being realistic about successes and barriers is key to developing best practices and making firm pedagogical advances in the field (McCormack et al., 2013).

It is also important to consider the role of the expert. The experts were chosen based on their role within the competition, how involved they were, what they could tell us about the process and their expertise. However, the children and the communities involved in the competition could also be considered experts. Most interviewees mentioned the desire to have access to data exploring village attitudes and continued use of the tablets, labelling this ‘valuable’ and ‘critical’. This data would be difficult to collect *post-hoc* following the competition due to the geographical remoteness and language barriers involved. Employing a mixed methods approach within the original research design would have been preferable.

Conclusion and Future Directions

This study has highlighted the importance and extent of community engagement in supporting different aspects of the implementation of an EdTech intervention with out-of-school children living in remote regions of Tanzania. This chapter has shown that app-based instruction can successfully support foundational learning with out-of-school children, but that EdTech alone seems insufficient to support the learning process as high levels of community support and engagement were required. This result challenges claims that EdTech pedagogy can promote autonomous learning without the need for adult support and demonstrates that for pedagogical practice with EdTech to be effective and sustainable in LMICs for out-of-school children, engagement with communities is essential.

These results have clear implications for practice as they emphasized the need for a thorough sensitisation process to ensure that villagers felt comfortable with the technology as it was being introduced into their community. Furthermore, engaging community members in the design process of educational apps might be beneficial to ensure content is culturally appropriate. Governments, implementers, and app developers should consider community members as critical partners in designing, deploying, and scaling educational technology interventions in remote, low-income settings to maximise overall success and sustainability. Given the global commitment to achieve the Sustainable Development Goal for Education by 2030 and ensure inclusive, equitable, quality education for all (UNDP, 2023), the insights gleaned from this study should be particularly informative to global educational organisations when implementing EdTech interventions with the world's most marginalised children.

Chapter 7 - General Discussion

This thesis had an overarching aim to identify potential predictors of foundational learning outcomes following a tablet-based EdTech intervention deployed to out-of-school children from rural communities in Tanzania. These were addressed at three levels of explanation, broadly answering the following questions to inform a new multi-level ecological model:

- What app-level features predict learning improvements?
- What child-level factors predict learning improvements?
- What community-level factors predict learning improvements?

As reported in Chapter 1, five initiatives have been identified that used app-based interventions to facilitate learning for out-of-school children (Gottwald et al., 2017; Orozco-Olvera & Rascon-Ramirez, 2023; Stubbé et al., 2016; Stubbé et al., 2017; XPRIZE, 2019). Results demonstrated that out-of-school children, including marginalised learners such as girls, can learn foundational literacy and numeracy skills using an app. However, significant methodological challenges were faced, learning was not truly autonomous, and little was known about the mechanisms that are driving the learning improvements. To address these research gaps, this thesis conducted a holistic mixed-method evaluation of the deployment of five learning apps in an out-of-school context, utilising the data, software, and insights from the GLXP competition to determine the potential predictors of foundational learning outcomes. The mechanisms through which out-of-school children may acquire foundational skills were investigated for the first time through four empirical studies (Chapters 2-6), each focusing on a different level of explanation. App features, gender and domain differences, child-level features, and community-level factors were evaluated to inform an ecological model specific to predicting learning attainment for marginalised out-of-school children.

To effectively understand how app features can support learning, they need to be evaluated when linked directly to learning outcomes for the first time. Chapter 3 (Comparative Judgement study) reported an experiment in which 41 non-expert participants compared the five finalist apps across 15 key pedagogical features to identify which characterises each app. Results indicated that the two apps that produced the greatest learning outcomes over the trial shared six pedagogical features—autonomous learning, motor skills, task structure, engagement, language demand and personalisation.

Chapter 4 (Domain and Gender App study) explored the effectiveness of different app features between genders, as this had never been done before in an out-of-school setting where numerous apps could be directly linked to learning outcomes and, thus, compared against one another. Machine learning methods and inferential statistics were used to identify which app features most predict girls' and boys' literacy and numeracy learning and whether there are any significant differences between domains and genders. Some app features, such as engagement and retrieval-based learning, were found to be broadly influential for learning. Five app features - curriculum links, personalisation, engagement, autonomous learning, and language demand- showed a differential influence across genders and domains.

Chapter 5 (Contextual Predictors study) used a machine learning regression approach to identify the contextual factors that most predict literacy and numeracy learning for the first time. Child and village-level features were leveraged using the competition's direct assessment and survey data and contextual covariates derived from open-source geospatial data. Reading habits, small family sizes, and low physical isolation were shown to be the most predictive factors for improved learning outcomes.

As the GLXP did not involve collecting qualitative data about the autonomous learning process throughout the competition, Chapter 6 (Expert Elicitation) aimed to qualitatively explore the broader implementational impact and challenges experienced. In this chapter, 14 key informants were interviewed about their experiences of the competition. Results demonstrated the importance of the cultural context in which the technology is deployed and the need for community support throughout the competition, challenging the extent to which children can learn autonomously with EdTech alone.

In Chapter 1, a new ecological model was introduced that attempted to depict potential predictors of learning improvement for out-of-school children using an EdTech intervention (see Figure 3). This provided a foundation upon which this thesis explored the predictors on three levels of explanation: app, child, and community. The results indicate that there are influencing mechanisms for app-based learning with out-of-school children at each of the different levels of explanation investigated in the thesis (e.g., app, child, community). This suggests that the learning process needs to be assessed at different levels to understand how out-of-school children learn using tablet technology. Evaluating just one level of influence is not sufficient to inform interventions with this demographic. This chapter will discuss the broader findings at each of these levels of explanation, considering the implications for progressing EdTech research and learning provisions targeting out-of-school children.

Level of explanation: App features

As discussed in Chapters 3 and 4, some app features were identified as being ‘neutral’ features, whereby they influenced children similarly, regardless of their gender or the domain being taught. The app features that were highly influential yet neutral for both genders and domains are retrieval-based learning, motor skills and task structure, demonstrating their broad relevance for learning improvements in out-of-school children. Therefore, developers should incorporate app features that have broad relevance and are likely to promote equitable learning for diverse demographics, focusing on features like retrieval-based learning, motor skills, and task structure. These elements emerged as highly influential regardless of gender or learning domain and should be prioritised to maximise reach and impact.

However, there are other app features that seem particularly important for one gender or a particular domain, and girls and boys were shown to differentiate in some of the app features that they found to be influential for learning. For example, in literacy, personalisation was found to be significantly more effective for girls learning attainment than boys, while the opposite was found for curriculum links. In numeracy, four features were significantly more influential for girls than boys: engagement, autonomous learning, language demand, and curriculum links. These features highlight the importance of tailoring apps to the targeted population, as there is clearly no ‘one-size-fits-all’ approach to app-based learning. The non-generic app features could be used strategically to promote equitable learning of marginalised groups, such as out-of-school girls who have notably struggled with access to education (Crompton et al., 2021c). However, incorporating app features that only benefit girls (e.g. personalisation for literacy) may risk promoting girls’ learning at the cost of boys being left behind and becoming a marginalised group themselves. This may suggest that it would be most beneficial to prioritise neutral features that promote learning for both genders to provide educational equality to all out-of-school learners.

For numeracy, it could be advised to prioritise engagement, autonomous learning, and language demand. These app features are technically features that target girls, as significant differences indicated that they benefitted girls’ numeracy achievement over boys. However, the three features all ranked in the top seven features across all four categories, demonstrating that they also have broad influence across both genders and domains. Therefore, prioritising these features could provide a pivotal balance between specifically targeting girls without excluding boys for numeracy app-based learning.

Clearly, governments and policymakers need to make important decisions using this evidence when considering app design and technological interventions, particularly considering how app features might be used to target specific learners. For the first time, this thesis studied the relative importance of app features in relation to learning outcomes. This provides the relevant decision-makers with clear statistical evidence about which key app features are most influential for learning improvements in out-of-school children following an app-based EdTech intervention.

Level of explanation: Child (context)

Chapter 5 showed that child-level factors are relatively useful in predicting children's baseline learning outcomes but are not very accurate in predicting learning improvements after an EdTech intervention. These findings have pivotal implications for implementation with out-of-school children, as they suggest that EdTech interventions could provide equal opportunities, whereby all children, irrespective of their background, have the opportunity to benefit. Furthermore, it implies that efforts and funds might be more effective if focused on the app design and intervention delivery rather than trying to pinpoint the child-based characteristics that make children more receptive to EdTech interventions.

However, as discussed in Chapter 5, many potential child-level features were not considered, such as nutrition, health, and developmental skills (e.g. cognitive, emotional, social). This is due to the restrictions of analysing secondary data, as the XPRIZE team did not collect this child-level information when conducting the contextual surveys with the children and their caregivers. Therefore, while some child-level factors were considered, further research is necessitated before concluding that community-level contextual factors are more important for predicting children's improvements following an EdTech intervention.

One key child-level characteristic that was not measured was whether the children had SEND or not. However, as outlined in Chapter 1, children with SEND are likely to be out of school and struggle to learn with EdTech at the same rate as their peers (Pitchford et al., 2018). Therefore, binary cut-off points were implemented using histograms of the EGRA and EGMA raw improvement scores in Chapter 4 to identify which children were considered low-achievers (or 'non-responders') at the end of an EdTech intervention. Out of 2041 children, 860 were considered low-achievers ('non-responders') for literacy improvement, and 467 were considered low-achievers for numeracy. This indicates that, as anticipated, there was likely a high proportion of children with SEND within the competition sample.

Novel data analytic methods (using classification models combined with SHAP analysis) were used in Chapter 4 to determine whether different app features can predict whether a child will be a low- or high-achiever. The fifteen app features could predict low or high achievement with an accuracy rate of 55.3-60.7% (across gender and domains), with marginally better accuracy for girls than boys. Future research could build on the regression research introduced in Chapter 5, using the binary split for low- and high-achievers and classification models. This will help determine whether there are child and community-level factors that can predict whether a child will be a non-responder to an EdTech intervention. This could facilitate the exploration of differences between non-responders and responders, helping educators to understand what characteristics may help identify children with SEND in an out-of-school population.

Level of explanation: Community (context)

Infrastructure

At the community level of explanation, Chapter 5 demonstrated that the village-level factors predominantly predicted learning improvements after an EdTech intervention. This demonstrates the importance of infrastructure and the children's environmental context when implementing EdTech solutions in remote communities with out-of-school children.

Proximity to local services and amenities, such as health services and police stations, was particularly important. Living close to a health centre or pharmacy may be predictive of learning as higher access to and use of services that prevent and treat disease positively influences health (WHO, 2017). This is significant in low-income countries like Tanzania, where diseases are rife among school-aged children, with an average malaria prevalence of 21.6% nationally (Chacky et al., 2018). Research has suggested that improving children's health could improve educational outcomes in developing countries, which could indicate why living close to health services was predictive of learning outcomes in this research (Glewwe, 2007). However, interventions in low-income countries that have provided children with preventative deworming medication or treatment for malaria have produced mixed results for educational outcomes (Ganimian & Murnane, 2016). Therefore, this complex interplay between health and education needs to be investigated further to establish whether improving out-of-school children's health could facilitate learning with EdTech solutions.

It must also be considered why proximity to a police station was predictive of learning improvements following the EdTech intervention. One reason may be the impact that proximity to the police has on Female Genital Mutilation (FGM) practices. FGM is a prominent issue for young girls in low-income countries in Africa, the Middle East and Asia, affecting more than 200 million women and girls (WHO, 2024). While this potential relationship has not yet been quantitatively investigated, qualitative research in Kenya exploring what is likely to motivate people to obey the FGM criminalisation laws cites distance to police stations or courts as a contributing factor (Meroka-Mutua et al., 2020). FGM has been shown to be a barrier to traditional education for girls, as it is often a prerequisite for early marriage, and the procedure itself can lead to pain, distress, and other health complications that prevent consistent school attendance (Orchid Project, 2021; The Commonwealth, 2014). These identified barriers of FGM further highlight the benefits of an EdTech intervention compared to traditional school attendance. Tablet-based technology can be flexibly used and does not require travelling to a school when unwell, enabling higher access for girls in rural, low-income communities. Ultimately, findings from this thesis suggest that proximity to a police station may reduce FGM practices, which in turn improves educational attainment for girls, although this interpretation warrants careful further investigation. This positive influence of proximity to key services indicates that all educational interventions may benefit from being more socially and physically connected, not just EdTech.

While community-level infrastructure, like proximity to health and law enforcement, was shown to be an important contextual factor linked to learning improvements, this should be interpreted cautiously due to the exploratory nature of the research. Further research is warranted to investigate how different types of infrastructure may influence learning in diverse low-income settings. This will help identify which infrastructure investments are most critical to enhancing learning outcomes within specific rural communities, as the current results do not offer a straightforward solution. For example, EdTech initiatives are presented as alternative learning provisions for children who live in remote locations and thus cannot easily access schools (XPRIZE, 2019). However, the current research suggests that if the communities are *too* physically isolated from amenities, the EdTech solution may not be as effective as for more centrally located children.

As mentioned in Chapter 1, a prominent infrastructural challenge for implementing EdTech in low-income countries is the cost. Digitally transforming education provision is a costly, ongoing investment that may be beyond the reach of the lowest-income countries (Antoninis, 2023). A review of EdTech provision in developing countries explored the cost of the hardware in existing initiatives, citing costs ranging from $150-200 per tablet or laptop unit (Rodriguez-Segura, 2022). However, many programmes did not report the cost, or only reported the cost per child without providing details about what this cost encompassed. This highlights a lack of transparency when reporting EdTech initiatives, which may impede decision-makers when trying to assess whether a learning solution is feasible in their setting. Furthermore, onebillion can provide their customised Android tablet, ‘onetab’, at a cost of $50 per unit when delivered at scale (minimum 1000 units), demonstrating that the costs of tablets can be much lower depending on the hardware chosen (onebillion, 2020). The lifespan of the tablet should span multiple years, further improving the cost for governments to provide tablets to children for foundational learning (onebillion, 2022).

Despite hardware costs lowering over recent years, the true cost of an EdTech initiative is often much higher than simply buying the hardware, even when implemented in an out-of-school setting. Piloting, software, sensitisation processes, labour, and the initial set-up of infrastructure all incur additional costs that need to be incorporated into an EdTech business model, even when the children are learning autonomously (Chuang et al., 2021). There are also substantial energy usage and connectivity costs if not using tablets and software like those in the GLXP, which were solar-powered and functioned without the internet (Chuang et al., 2021; XPRIZE, 2019). Furthermore, when implementing an EdTech initiative with out-of-school children, a supply chain needs to be costed and planned for the replacement of broken devices. Broken tablets were a common issue faced in the XPRIZE competition (Chapter 6), and lack of ongoing distribution may threaten the sustainability of tablet-based learning.

Another point to consider when costing an EdTech programme is the complementary investments of extra (optional) infrastructure when trying to maximise the effectiveness for both cost and learning improvements (Chuang et al., 2021). As this thesis has demonstrated the importance of community support or out-of-school learning with EdTech (Chapter 6), additional infrastructure may include setting up learning centres where volunteers or staff can provide facilitated learning sessions, as seen in existing initiatives (e.g. Stubbé et al., 2016; Education Development Center, 2020). However, a sandbox event was held where the EdTech Hub and onebillion partnered to explore the potential of scaling their Unlocking Talent 'onetab' initiative nationally in Malawi (Longley, 2021). They explored three different implementation methods for their EdTech programme: in learning centres, in school and at home. It was found that using learning centres was significantly more expensive at $30 (per child per year) than the other two options ($7 in school, $6 at home), deeming it too expensive to scale nationally (Longley, 2021). However, the learning centres in this cost example were not set up for out-of-school children. They were part of an 'extraction model' where facilities were built on the school campus, and students were removed from classrooms to learn with the tablets. Therefore, it would be beneficial to investigate this comparison for out-of-school children, assessing the cost difference between learning in the home versus in a built-for-purpose learning centre. While self-led learning is a promising area for cost-effectiveness, the importance of community support should not be underestimated (Rodriguez-Segura, 2022). Important decisions would need to be made by weighing up the costs and benefits of each option before deciding whether to implement an autonomous or community-supported intervention.

As previously mentioned, proximity to services such as pharmacies is highly influential for children's foundational learning improvements. The importance of this extra infrastructure means that governments have additional decisions to make when aiming to maximise the effectiveness of EdTech implementation. Setting up localised facilities, such as pharmacies or police stations, could be incorporated into the implementation process but would likely be costly and not within limited educational budgets for low-income countries.

Furthermore, if governments are unable to fund local schools in the rural areas this thesis is targeting, it is unlikely that building extra pharmacies and local amenities is a feasible possibility. However, it could be argued that as EdTech interventions can effectively improve learning outcomes for out-of-school children (as in the GLXP), governments might consider prioritising building pharmacies, police stations and other key services over putting schools in these remote locations. One reason is that school infrastructure and teaching provision are often poor-quality in developing countries, leading to stagnation in learning improvement for children (Fomba et al., 2022; UNESCO, 2019). Building pharmacies and police stations is multi-purpose, as their proximity to remote communities may promote safety and health and facilitate tablet-based learning. As mentioned in Chapter 1, governments also prefer to invest in tangible assets, such as buildings, so they might be more amenable to this solution, particularly as healthcare and safety provision benefits the whole community (Passey et al., 2016; Kaguo, 2011). To minimise the costs, facilities could be positioned within close proximity to multiple communities where possible, similar to how the XPRIZE competition placed their solar-powered charging stations. Ultimately, this research has shown that there are prominent cost implications for implementing an EdTech programme within remote, low-income communities. This leaves careful decisions to be made by the local and national governments about whether the results are worth the infrastructural challenges, with cost remaining one of the most difficult barriers to overcome.

This thesis offered unique insights into the impact of community-level factors, including engineered geospatial features, on learning outcomes following an EdTech intervention. Considering contextual factors is crucial to holistically evaluate interventions and ensure contextually informed programme design (Nicolai et al., 2023). However, this relationship between community-level contextual features and learning outcomes has never been quantitatively explored before in EdTech research, not even in a school setting. At the classroom level, Outhwaite et al. (2019) explored implementation themes that predicted learning outcomes following a Maths-based EdTech intervention and found that 'established routine' significantly predicted learning outcomes. A recent review exploring COVID-19 research to investigate the effectiveness of implementation contexts highlighted contextual issues faced for out-of-school EdTech learning, such as stable electricity, connectivity, and appropriateness of provided resources (Nicolai et al., 2023). However, as previously identified, emergency COVID-19 learning provision is distinct from longer-term out-of-school provision. Furthermore, contextual elements were considered important based on measures of parental confidence, learner preferences and engagement (Nicolai et al., 2023).

The current research builds on this research by quantitatively linking relevant contextual factors to learning improvements following an EdTech intervention using novel data analytic methods. As these factors were shown to be highly influential, it would be valuable to employ the same geospatial and machine learning methods in school-based settings to determine the key contextual predictors of learning improvements in a traditional setting. For example, in Tanzania, there is a national exam called the Primary School Leaving Examination, which children must pass to progress to secondary education (UNESCO, 2024b). The exam results could be used to determine what predicts learning outcomes in a school environment. Alongside the community-level factors utilised in the current research (e.g. proximity to local services, number of buildings and roads in the villages), the influence of school-level factors could also be studied for the first time, such as the size of the school/classes and the number of teachers. Analysing the community-level factors for school-based learning will help researchers and local governments understand what contextual factors are specifically influential for EdTech interventions and which are broadly influential for learning in these low-income contexts. Building on this research further with the school-level contextual predictors (and existing classroom-level research) could inform decision-making by illuminating the organisational areas that should be prioritised when governments are aiming to improve the quality of existing schools.

Implementation

At the community level, implementation was explored in Chapter 6 using key informants from the competition. One of the key themes placed emphasis on the extent to which children's families, peers, and wider communities were believed to substantially support their learning experience with scaffolding, support, and motivation. This suggests that a responsive, helpful community may be pivotal to children learning with the EdTech in an out-of-school setting. However, Chapter 5 analysed features based on whether the children perceived themselves to receive support from their parents, siblings, and the village Mamas, and did not find them predictive of children's learning improvements. Although this seems to undermine the importance of communities, these features were engineered based on questions from the survey asking if an individual supported their tablet-based learning, with a yes/no answer. The data was not on a scale, with levels of support assessed (low-high), and thus, it is somewhat insufficient to establish in what capacity people supported the children's learning throughout the competition. Therefore, anecdotal evidence from key informants of the competition may be a more reliable source of knowledge about the importance of communities during the implementation process. A more objective measure of community involvement and caregiver support may be necessary to reliably determine the role of the community in learning with EdTech in out-of-school settings.

The identified importance of communities has pivotal implications for the success of EdTech interventions with out-of-school children. The consistent support provided by communities threatens the notion that children can learn autonomously with tablet-based technology and a well-designed app. This suggests that to implement a successful EdTech intervention, the technology and software cannot just be deployed to children without organising consistent support from the surrounding communities. As identified in Chapter 1, most existing out-of-school initiatives in low-income countries already incorporate *some* level of community-based support into their implementation programme, whether that is just technical support, includes motivational feedback, or the facilitation of regular learning sessions (i.e. informal teaching support; Stubbé et al., 2016; Brezael et al., 2016). The current research provides evidence that community involvement is important for successful EdTech interventions but highlights that further research is needed to identify which elements of the support are most beneficial. While community support has been shown to be important, EdTech interventions should involve an initial assessment of community receptiveness and capacity for support. Pragmatically, future research might explore scalable community support models, where high-touch engagement is initially provided but gradually tapered as community familiarity and self-efficacy with the technology increase. The importance of community-based scaffolding also implies that it is vital for local governments to establish how receptive a community will be to an intervention before implementing it. Community attitudes could influence a programme's overall success (explored further under macro-level factors), so they may inform key decisions around which communities to choose for implementation. A flexible, phased approach to community support might be necessary, allowing interventions to adapt to local needs and resources as they develop.

Levels of explanation not explored in this thesis

As highlighted, this thesis explored the app-level, child-level, and community-level factors within an ecological model that may impact the foundational learning improvements for out-of-school children using EdTech. However, two levels of the ecological model, macro- and chrono-level factors, were not explored within the scope of this thesis. This section highlights areas of priority within these levels of explanation and suggests future research that could address these research gaps to build a fully comprehensive holistic view of out-of-school EdTech provision.

Level of explanation: Macro

One level of explanation that was not specifically studied in this thesis is the macro-level factors, such as cultural attitudes and governmental policies. However, Chapter 6 (Expert Elicitation) explored the implementation experiences of the GLXP, and two of the resulting themes highlighted valuable points around the cultural experience within communities.

One theme centred around the novelty of the technology, which caused initial confusion and distrust within the communities. This highlights the importance of a careful sensitization process for communities. This has implications for the implementation of EdTech interventions in rural communities in Africa, as it may suggest that Western implementers (as in the GLXP) should partner up with local researchers and educators to develop and deliver sensitization processes. This can help ensure the sensitization process is contextually appropriate and provides communities with a safe space to share their concerns and questions.

Another theme generated in Chapter 6 was the importance of knowing and respecting the specific cultural context of the location in which the intervention will be implemented. This reinforces the need for a well-planned sensitisation process, but also challenges the scalability potential of apps across different countries and cultures. However, existing apps using tablet technology have been shown to be successful at improving learning outcomes in different settings and countries. For example, onebillion have shown success in improving maths outcomes using their specialised software and tablet-based technology in the UK, Malawi, Brazil, Sierra Leone, and Tanzania (Outhwaite et al., 2019a; Pitchford & Outhwaite, 2016a; Outhwaite et al., 2020; Lurvink & Pitchford, 2023; XPRIZE, 2019). This suggests that not all elements of the app need to be engineered to be culturally specific to the children's setting. As suggested in Chapter 6, the findings underscore the need for EdTech developers to consider cultural context during implementation (Huntington et al., 2023b). Partnering with local researchers and educators to design and deliver sensitisation processes can help ensure that EdTech programmes are culturally appropriate and aligned with local values. Additionally, developers should explore how collaborative learning features can be adapted to support collectivist cultures, making them more effective for learners in these settings (Nicolai et al., 2023).

While the expert elicitation themes touched upon the cultural attitudes and norms during the competition, this macro-level factor remains underexplored throughout this thesis. A limitation of the GLXP competition is that no follow-up research was conducted once it was finished and the winners announced, as that was not within the scope of the competition's goals and timeframe. The nature of funded competition-based research means that the XPRIZE Foundation had its next project to work on following the conclusion of the trial. A follow-up study to evaluate long-term use and attitudes towards the tablets would have been beneficial, especially as post-intervention research studies can inform long-term implementation and sustainability and provide insights into the general acceptability of RCTs within the target audience (Pegrum et al., 2013). As part of this PhD studentship, it was originally planned that a follow-up study would be conducted in the Tanzanian villages. The study would have used a combination of observational, survey, and interview techniques to assess the prolonged use of the tablets once the competition had finished and the general acceptability of the software (from children, parents, and the wider community members) as an alternative solution where traditional schooling is not available. In 2020, a global COVID-19 pandemic led to a national lockdown in the UK. It restricted travel to Tanzania from 2020-2022 to reduce the spread of the virus, preventing the follow-up study from being carried out as planned. The potential of a follow-up study was reconsidered during the final year of study in 2023, but XPRIZE would not be able to facilitate the research visit as previously planned due to staffing changes, so a follow-up visit was not possible within the timeframe of this studentship. However, future EdTech trials with this population would likely benefit from a follow-up study to evaluate direct community attitudes and see if interest and use were sustained over time. This should be planned into the initial research protocol so it can be implemented in a timely and organised manner.

Level of explanation: Chrono

Another level of explanation not explored in this thesis was the chrono-level factors, which are factors measured over time. The GLXP competition was focused on immediate learning gains at the endline of the 15-month EdTech intervention. No sustained learning gains were tested with longitudinal data, so this could not be analysed in the thesis. A second result of the technology being novel was that children were interested and excited to learn with the tablets, which was believed to drive engagement throughout the competition. However, it remains unclear whether this interest and motivation will be sustained long-term when the novelty of new technology has worn off. Whilst the effectiveness of EdTech interventions has been established through RCTs, some results have shown that positive gains often start to diminish over time, although there are early signs of sustained success in some school-based research (Tauson & Stannard, 2018; Outhwaite et al., 2017). Additional research is needed to identify and address factors influencing the sustainability of effects in EdTech interventions and to employ follow-up or longitudinal methods to implement interventions that allow long-term outcomes to be tracked effectively (Walton, 2018). Determining whether interest and learning gains can be sustained is crucial to inform policy-based recommendations, as governments have limited budgets and thus must invest in learning solutions with the most long-term promise and impact.

Another chrono-level factor that was not explored in this thesis is app updates of the learning software used. In the competition, each finalist team was allowed to update their app software once at a time of their choosing, changing any elements they wished to (XPRIZE, 2019). These changes were made based on any bugs identified and the weekly usage data taken from the villages, in which time spent in activities within the apps was visible, but achievement statistics were not. It would have been useful to conduct midline assessments after each new version of the app is implemented to evaluate whether the updates had an impact on learning or motivation, although the nature of the competition means this would not have been feasible in the current intervention. This could help app developers evaluate the influence of any big changes made to the software, such as the impact of specific app features (e.g. increasing feedback within maths games).

Another factor that was not studied is the children's development over time. This could include their academic progress and engagement levels; these could be monitored using midline and longitudinal assessments, as suggested for app updates. This would identify if children made quicker academic progress as they build up their foundational skills from a starting point of illiteracy and innumeracy. Building on this, it was shown in Chapter 2 (Case study) that 276 (13.52%) of the children that participated were attending school by the time of the endline assessment (March 2019). These were all out-of-school children at the start of the competition period, as necessary for the aim of the trial. While an increase in children having access to and attending school is highly positive, it does have important implications for the data. There is no record of when the children began attending school, and introducing such changes to the trial could confound the results, as it is impossible to know which learning outcomes should be attributed to the intervention and which potentially were a result of their school experiences. Results from Chapter 5 (Contextual Predictors) indicate that prior school attendance is an influential predictor of learning outcomes at baseline (prior to any learning intervention), which may suggest that unplanned school attendance could also have impacted the positive learning outcomes following the EdTech intervention. The timing may suggest that the uptake in school attendance directly results from participation in the competition (i.e.. because of an increase in motivation or change in parental attitudes), but there is no evidence to confirm whether this is true. Building in the suggested midline assessment may help combat the threat that changes such as newly attending school have on the quality of the data collected for an out-of-school EdTech initiative, as it allows these changes to be carefully tracked and incorporated into analyses.

Another aspect of children's development not studied is their socio-emotional development, which is their ability to learn fundamental skills such as self-concept, self-efficacy, the ability to express themselves, and the ability to form and sustain relationships with others (Rymanowicz et al., 2020). Socio-emotional data was collected by UNESCO at both baseline and endline in the competition, assessing the children's relationships, emotions, and social skills (XPRIZE, 2019). However, this data was not available for analysis in this thesis. If it was, social-emotional data could have been used to create pivotal child-level features for the regression analysis that identified influential contextual predictors of learning in Chapter 5. Furthermore, as the socio-emotional data was collected at both baseline and endline, it would have been possible to evaluate developmental changes over time and assess whether an EdTech learning intervention bolsters socio-emotional skills. Previous research has shown that EdTech interventions can enhance socio-emotional skills in young children (Vegas et al., 2019; Tauson & Stannard, 2018). If this could be shown in out-of-school children in low-income countries, it would further illuminate the benefits of EdTech solutions, strengthening the argument presented to governments about the advantages of EdTech implementation.

Reflections on the chosen outcome measures

While the XPRIZE team chose the outcome measures used for the Global Learning XPRIZE, it is pivotal to critically review their use and potential limitations within the context of the current research. The Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA) are standardised methods used to measure foundational skills in literacy and numeracy, which were the desired outcomes of the GLXP competition (Gochyyev et al., 2019). The two assessments primarily assess basic skills, such as phonological awareness, letter and number recognition, and basic arithmetic. They are unable to capture broader competencies, such as problem-solving or critical thinking, which may limit their ability to measure the full range of skills that can be developed through interactive EdTech and are also crucial to children's early literacy and numeracy development (Ochea & Hernandez, 2023).

Another potential limitation of using the EGRA and EGMA measurements is that many children scored extremely low scores following the intervention (i.e. a score of between 0-10), indicating that there may have been a floor effect. This could mean that the EGRA and EGMA were too difficult for the children, and potentially not sensitive enough to capture the very early stages of skill acquisition and accurately reflect their learning gains (Gochyyev et al., 2019). This may be a more prominent issue in the current research, as although designed for early grade learning assessment, the EGRA and EGMA are not designed with the specific context of out-of-school children in mind, who may be older or at different developmental stages. If the adaptation process is insufficient, the assessments may not capture children's learning in a way that is meaningful to their day-to-day experiences and lives, particularly in a context like rural Tanzania where multiple dialects are spoken and the cultural environment can vary hugely (Huntington et al., 2023a; Ochea & Hernandez, 2023).

To better capture learning outcomes, future research could complement EGRA and EGMA by adding broader assessment tools that measure cognitive, digital, and problem-solving skills. For example, digital literacy tools like the Multiple Indicator Cluster Survey (MICS)'s Foundational Learning Skills could provide a more holistic view of children's learning outcomes. This would be particularly useful when testing out-of-school children, as their learning trajectory may be less linear than for those in traditional schooling, and thus a more nuanced approach may be necessary to adequately capture learning progress (Gochyyev et al., 2019; Ochea & Hernandez, 2023).

If an alternative test was needed to avoid the floor effect evidenced in the current research, the Annual Status of Education Report (ASER) could be a suitable replacement. The ASER is designed to measure the most basic literacy and numeracy skills, and is widely used in rural, low-income settings in countries across Sub-Saharan Africa, India and Pakistan, making it contextually suitable (Banerji et al., 2013; Banerji & Chavan, 2021). Its simplicity and scalability make it a highly effective tool for assessing learning in low-resource, informal contexts, and it has proven applicability across various education systems and contexts (Banerji et al., 2013; Banerji & Chavan, 2021).

In summary, although EGRA and EGMA was a sensible choice for the GLXP, incorporating more sensitive and culturally relevant assessments alongside EGRA and EGMA may better capture the full scope of children's learning progress in low-resource, out-of-school contexts.

Theoretical model of EdTech learning for out-of-school children

As evidenced throughout this chapter, investigating an EdTech intervention with out-of-school children through the lens of a new ecological model has illuminated that multiple mechanisms influence foundational learning improvements at each level of explanation. This suggests that to understand an EdTech intervention with out-of-school children, you cannot focus on one singular level of explanation. Figure 11 incorporates the findings from this thesis into the initial model introduced in Chapter 1, creating an updated ecological model depicting the driving mechanisms behind learning with EdTech for out-of-school children. Influential factors were added to the model, while any features found not to positively impact learning outcomes for out-of-school children were removed (e.g. feedback as an app feature).

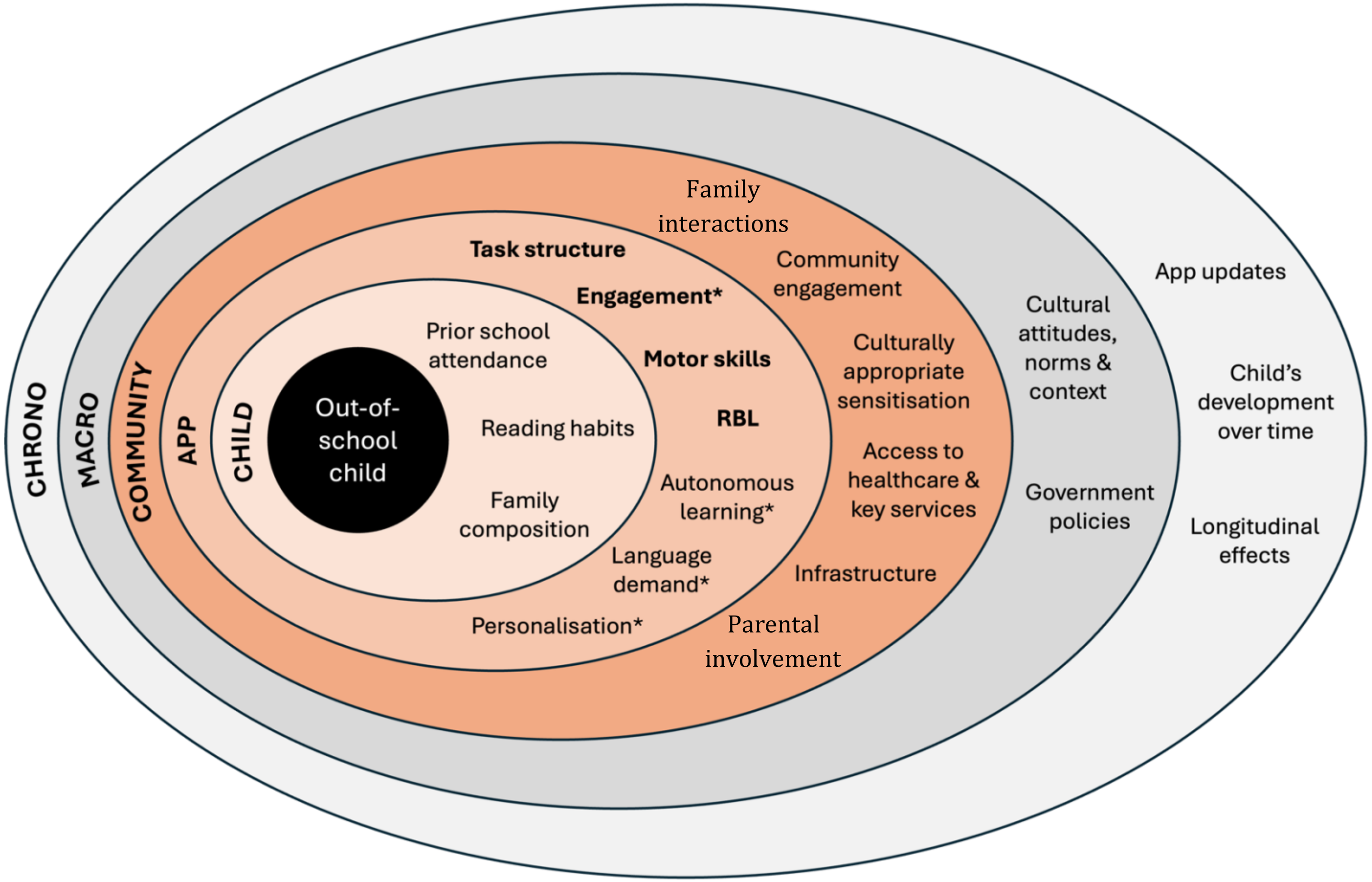
This builds on the work of Outhwaite (2019), who created an ecological model for Maths app-based learning in school settings and then embedded multiple findings from studies conducted with the 'onebillion' app within the framework. While the school-based model had primarily different focus areas, comparisons can be made between the findings embedded within the two models to identify similarities and differences between school-based and out-of-school learning with EdTech. Outhwaite (2019) found that learning gains are not associated with SES. While child SES was not directly studied within this thesis, caregiver profession was used as a proxy for parental SES. It was found that out of all professions studied, unemployed caregivers were the biggest predictors of learning improvements (Chapter 5), which consolidates Outhwaite's findings. This suggests that across all environments, EdTech interventions may help all children learn regardless of their socioeconomic backgrounds, highlighting the potential for low-income children. Another child-level finding from Outhwaite's model was that progression in the apps is dependent on the child having sufficient proficiency in the language of instruction (Outhwaite, 2019). This thesis evaluated the importance of language from an app-level perspective rather than child-level. Providing accessibility via language demand was found to be highly influential across domains and genders, supporting Outhwaite's prior findings by emphasising the importance of language skills for all children. The research suggests that either apps need to be simple and clear enough for children with lower language proficiency, or the children need to have proficient language skills. In a low-income environment where language proficiency is likely to be low and children are more likely to have SEND (Moumen, 2023), incorporating simple language within learning apps is advisable to minimise barriers to access for EdTech-based learning.

Several key differences were also found between the two models, which may be because of the differing learning environments. One such difference is the fact that Outhwaite (2019) found that learning gains were not associated with child engagement. While child engagement was not directly measured in this thesis, the two winning apps of the GLXP greatly embodied ‘engagement in the learning process’ as an app feature, and it was arguably the most influential app feature for improving learning outcomes across both domains and genders. Therefore, this contrast may suggest that facilitating engagement with the app is more important for out-of-school children than children in a school environment. This is unsurprising, as out-of-school children are required to work more autonomously than children in schools, with less structure and support, so they rely on higher levels of engagement from the app itself to keep them interested. At the contextual level, in a school environment, Outhwaite (2019) found that teacher support and supervision was not associated with app-based learning gains. However, this thesis demonstrated that community support was found to be critical to the children’s learning success, and that caregivers and wider community members supported the learning process with motivational encouragement, peer-based learning opportunities, and providing a safe space to learn (Chapter 6). Comparing the two models has highlighted key differences in the influence of child and contextual factors on learning gains following an EdTech intervention in both school and out-of-school environments. This supports the development of a new model for out-of-school children, as it demonstrates that the different contexts in which EdTech is implemented makes a crucial difference to the holistic approach needed to facilitate learning gains with an EdTech intervention.

As suggested above, future research would be highly valuable to further build this model, particularly for the macro- and chrono-levels that were not formally evaluated in this thesis. This would be beneficial to create a fully holistic model for out-of-school children, making a critical contribution to educational policy and practice in low-income countries where additional methods of learning are required.

Figure 11

Thesis results embedded within the multi-level model for predictors of learning improvement for out-of-school children using an EdTech intervention



Note. Within the app level of the model, bold text indicates broadly influential features, and * indicates where an app feature is influential but interacts significantly differently across domains and genders. The grey parts of this model highlight the levels that were not studied within the scope of this thesis, so the factors hypothesised to be influential require further examination.

Value of mixed-methods approach

The use of a mixed-methods approach to examine the app, child, and community-level predictors throughout this thesis was valuable for several key reasons. Firstly, as mentioned in Chapter 2, the combination of different analytical methods - namely inferential statistics, data analytics and thematic analysis - allowed for a more robust exploration of predictors of out-of-school learning on three differing levels of explanation. Combining research methods was necessary to comprehensively answer the primary aim of this thesis. It helped to address the weaknesses of each individual approach and maximise what can be discovered with the provided data. For example, traditional statistics (e.g. linear regression) could have been used to identify app features that predict learning improvements, but the use of the Bradley-Terry model and SHAP analysis (Chapter 4) allowed for a more sophisticated analysis that identified the 'relative weights' or importance of each feature in relation to one another. It is the first time that the relative importance of educational app features has been analysed, and this knowledge will be extremely useful for app developers to prioritise pedagogical features when designing learning app software.

The most significant novelty of the research in this thesis is the use of data analytics and machine learning methods, as well as the ways these techniques were implemented, to evaluate EdTech interventions in out-of-school settings. As identified in Chapter 2, big data research is scarcely utilised in the EdTech field, with researchers yet to discover its true potential (Longley, 2022; EdTech Hub, 2022). Recent research has used machine learning methods in school-based research (without EdTech) to determine the key predictors of educational outcomes in areas such as literacy, maths, and language (e.g. in Vietnamese primary schools, Musso et al., 2020; and a review of university-based research, Ofori et al., 2020). These classification models are also used to identify which children are low or high achievers (as in Chapter 4), intended to create targeted learning interventions for the lowest achievers (Musso et al., 2020). However, the previous research does not evaluate what factors influence learning improvements over a period of time, but just their educational outcomes at the point of testing to create a prediction model that characterises high or low achievement. While this thesis created a similar baseline prediction model (Chapter 5), it also built on this research by evaluating what external factors influence learning improvement, presenting more comprehensive results that can be carefully studied to inform design for future EdTech interventions. For the first time, these prediction models were built for out-of-school children. This is pivotal as out-of-school learners are a distinct population, so school-based prediction models are unlikely to be generalisable and relevant for this demographic (Giraldo et al., 2021).

As part of the contextual predictors research in this thesis (Chapter 5), multiple features were engineered using open-source geospatial data files accessed from online repositories, such as the number of buildings and roads in the villages. Other geospatial features were created by using this data and online mapping services (e.g. OpenStreetMaps) to calculate distances from the villages to local amenities. Using geospatial data in this manner is a novel technique for educational research that showcases just a few possibilities of how online data can be manipulated and incorporated into EdTech evaluations. The increased availability of big data as EdTech distribution increases in low-income countries (e.g. tablet usage data; Longley, 2022) and the exponential development of AI-assisted mapping services (e.g. the Building Footprints project; Microsoft Bing, 2020) provide many possibilities and potential for the quality of educational research in coming years. Incorporating geospatial data with EdTech research should enhance the informed recommendations made for policy and practice, as it will mean that they are data-driven, dynamic, and relevant to the specific country or context they are targeting. It is crucial that these opportunities are capitalised on during this period of learning stagnation and crisis.

Conclusion

This thesis utilised existing EdTech interventions - through the GLXP competition - to explore app, child, and community-level predictors of foundational learning improvements for out-of-school children in rural villages in Tanzania. A novel multi-methodological approach was adopted to examine predictors at multiple levels of explanation and determine how relevant mechanisms influence learning with tablet-based technology.

Six key pedagogical app features were found to be influential for learning, as they were embodied strongly across the two winning apps of the XPRIZE competition. Three of the apps were centred around the software's accessibility, demonstrating that this is important for all children, not just those with SEND. Some app features were influential for all learners across domains, while others were differently influential depending on gender and domain, highlighting that there is no one-size-fits-all solution to app design for out-of-school children.

To successfully implement a learning app, this thesis has shown that it is important to consider child and community-level contextual factors. For the first time, it has been demonstrated that community-level predictors are more influential than child-level features for predicting learning improvements, showing the importance of feeling physically and socially connected to the learning process. Discussion around the implementation of the GLXP highlighted how important community-based support is to the success of an EdTech intervention, which questions the feasibility of purely autonomous learning with tablet-based technology. Careful sensitisation processes could be used to explore a middle ground that balances community-based support with a level of autonomous learning in remote communities.

Ultimately, this thesis explored predictors of learning on three different levels of explanation before embedding the results into a new ecological model for out-of-school children. The model helps to enhance the limited theoretical understanding of how out-of-school children learn with tablet-based technology. It has led to several recommendations for future research to build on these findings further across different contexts and populations. The results of this thesis highlight the potential of tablet-based technology as an alternative learning provision for out-of-school children. If influential mechanisms are carefully incorporated into the design and implementation of EdTech interventions, they may support the most marginalised learners, such as girls and children with SEND. This has pivotal implications for educational policy and practice in low-income countries when leveraging EdTech solutions to try and address the global learning crisis.

References

Abbate, J. (1999). Getting small: a short history of the personal computer. *Proceedings of the IEEE, 87*(9), 1695-1698. https://doi.org/10.1109/5.784256

Adam, T., Haßler, B., & Cruickshank, H. (2016). One laptop per child Rwanda: Enabling factors and barriers. *Empowering the 21st Century Learner*, 184–195. https://aa-rf.org/wp-content/uploads/2023/06/saicet-2016-proceedings-tech.pdf#page=194

Admiraal, W., Huizenga, J., Heemskerk, I., Kuiper, E., Volman, M., & ten Dam, G. (2014). Gender-inclusive game-based learning in secondary education. *International Journal of Inclusive Education*, *18*(11), 1208–1218. https://doi.org/10.1080/13603116.2014.885592

Adukia, A., & Evans, D. (2023, October 12). *Most Out-of-School Children Are in Rural Areas. Education Systems Must Serve Them Better.* Center for Global Development. https://www.cgdev.org/blog/most-out-school-children-are-rural-areas-education-systems-must-serve-them-better

Afoakwah, E., Carballo, F., Caro, A., D'Cunha, S., Dobrowolski, S., & Fallon, A. (2021). *Dialling up Learning: Testing the impact of delivering educational content via Interactive Voice Recognition to students and teachers in Ghana* [Working Paper 39]. https://doi.org/10.53832/edtechhub.0051

Ahmad, I. (2019). Performance of Classifiers on Noisy-Labeled Training Data: An Empirical Study on Handwritten Digit Classification Task. *Advances in Computational Intelligence; Lecture Notes in Computer Science*, *11507*, 414–425. Springer. https://doi.org/10.1007/978-3-030-20518-8_35

Al-Azawi, R., Al-Faliti, F., & Al-Blushi, M. (2016). Educational gamification vs. game-based learning: Comparative study. *International Journal of Innovation, Management and Technology*, *7*(4), 132-136.

Al-Samarrai, S., & Tamagnan, M. E. (2019). *Gender Equity and Fee-Free Basic Education in Tanzania Summary*. World Bank. https://documents1.worldbank.org/curated/en/356111553606355438/pdf/Gender-Equity-and-Fee-Free-Basic-Education-in-Tanzania.pdf

Alam, S. S., & Dubé, A. K. (2022). Theoretically driven educational app design: the creation of a mathematics app. *Educational Technology Research and Development*, *70*, 1305–1327. https://doi.org/10.1007/s11423-022-10109-9

Alemán de la Garza, L., Anichini, A., Antal, P., Beaune, A., Crompton, H., & Tsinakos, A. (2019). *Rethinking pedagogy. Exploring the Potential of Digital Technology in Achieving Quality Education.* UNESCO MGIEP.

Allen, M. (2017). Field Experiments. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods*. SAGE.

Allen, R., Burgess, S., & McKenna, L. (2014). *School performance and parental choice of school: secondary data analysis: Research report*. Department for Education. https://assets.publishing.service.gov.uk/media/5a7c1bebed915d1c30daa9e1/RR310_-_School_performance_and_parental_choice_of_school.pdf

Allier-Gagneur, Z., & Coflan, C. M. (2020). *Your Questions Answered: Using Technology to Support Gender Equity, Social Inclusion and Out-Of-School Learning* (Helpdesk Response No. 14). EdTech Hub. https://doi.org/10.53832/edtechhub.0025

Allison, C. (2023). *Guidance note on using implementation research in education*. USAID. https://www.edu-links.org/resources/guidance-note-using-implementation-research-education

Allison, K. (2019). *An evaluation of tablet-based apps for maths learning in young children with Down's Syndrome* [Unpublished doctoral dissertation, University of Nottingham].

Amenya, D., Fitzpatrick, R., Njeri, M. E., Naylor, R., Page, E., & Riggall, A. (2021). *The Power of Girls' Reading Camps: Exploring the impact of radio lessons, peer learning and targeted paper-based resources on girls' remote learning in Kenya*. EdTech Hub. https://doi.org/10.5281/zenodo.4923094

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly, 27*(2), 231–244. https://doi.org/10.1016/j.ecresq.2011.08.003

Andrabi, T., Das, J., & Khwaja, A. I. (2012). What Did You Do All Day?: Maternal Education and Child Outcomes. *Journal of Human Resources*, *47*(4), 873–912. https://doi.org/10.1353/jhr.2012.0029

Antoninis, M. (2023). *The cost of digital transformation is well beyond poor countries' reach*. Global Partnership for Education. https://www.globalpartnership.org/blog/cost-digital-transformation-well-beyond-poor-countries-reach

Apple. (2007, January 9). *Apple Reinvents the Phone with iPhone*. Apple Newsroom. https://www.apple.com/uk/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/

Araya, R., Arias Ortiz, E., Bottan, N. L., & Cristia, J. P. (2019). *Does Gamification in Education Work?: Experimental Evidence from Chile.* Inter-American Development Bank. https://doi.org/10.18235/0001777

Ashlee, A., Clericetti, G., & Mitchell, J. B. (2020). *Refugee Education: A Rapid Evidence Review.* EdTech Hub. https://doi.org/10.5281/zenodo.4557019

Aslam, M., & Kingdon, G. G. (2012). Parental Education and Child Health—Understanding the Pathways of Impact in Pakistan. *World Development*, *40*(10), 2014–2032. https://doi.org/10.1016/j.worlddev.2012.05.007

Awinia, C. (2019). Free Basic Education and Gender Disparities in Tanzania: An Empirical Assessment of Challenges and Policy Options. *Huria: Journal of the Open University of Tanzania*, *26*(2).

Ayala, G. X., & Elder, J. P. (2011). Qualitative methods to ensure acceptability of behavioral and social interventions to the target population. *Journal of Public Health Dentistry*, *71*, S69–S79. https://doi.org/10.1111/j.1752-7325.2011.00241.x

Ayiro, L. P., & Sang, J. K. (2016). Provision of Education to the "Hard to Reach" Amidst Discontinuity in Nomadic Communities in Kenya. *FIRE: Forum for International Research in Education*, *3*(3). https://doi.org/10.18275/fire201603031070

Azevedo, J. P., Hasan, A., Goldemberg, D., Geven, K., & Iqbal, S. A. (2021). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer*, *36*(1), 1-40. https://doi.org/10.1093/wbro/lkab003

Bacher-Hicks, A., & Goodman, J. (2021). The Covid-19 Pandemic Is a Lousy Natural Experiment for Studying the Effects of Online Learning: Focus, instead, on measuring the overall effects of the pandemic itself. *Education Next*, *21*(4), 38–43.

Badilla-Quintana, M. G., Ramírez-Peña, G., Sandoval-Henríquez, F. J., Sáez-Delgado, F., & Gómez-Franco, L. (2022). The use of mobile technology in the development of cognitive skills of high school students with special educational needs. *Aula Abierta*, *51*(3), 227–236. https://doi.org/10.17811/rifie.51.3.2022.227-236

Banerji, R., & Chavan, M. (2021). *The Annual Status of Education Report (ASER) and the challenge of universal access to schooling in India*. In P. McMillan & K. L. Rogers (Eds.), Learning and education for a better world: The role of social movements (pp. 171-187). Springer.

Banerji, R., Goyal, S., & Kannan, K. (2013). *Annual Status of Education Report (ASER) 2012: Rural*. ASER Centre.

Bardack, S., Lopez, C., Levesque, K., Chigeda, A., & Winiko, S. (2023). An exploratory analysis of divergent patterns in reading progression during a tablet-based literacy program. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.983349

Barnes, E., & Puccioni, J. (2017). Shared book reading and preschool children's academic achievement: Evidence from the Early Childhood Longitudinal Study-Birth cohort. *Infant and Child Development*, *26*(6), e2035. https://doi.org/10.1002/icd.2035

Baytak, A., Tarman, B., & Ayas, C. (2011). Experiencing technology integration in education: children's perceptions. *International Electronic Journal of Elementary Education*, *3*(2), 139-151.

Beeharry, G. (2021). The pathway to progress on SDG 4 requires the global education architecture to focus on foundational learning and to hold ourselves accountable for achieving it. *International Journal of Educational Development*, *82*, 102375. https://doi.org/10.1016/j.ijedudev.2021.102375

Beffel, J. H., Gerde, H. K., & Nuttall, A. K. (2021). Siblings and Interventions: How Siblings Influence Development and Why Practitioners Should Consider Including Them in Interventions. *Early Childhood Education Journal*, 1–10. https://doi.org/10.1007/s10643-021-01273-3

Behrman, J. R., Murphy, A., Quisumbing, A. R., & Yount, K. (2009). *Are Returns to Mothers' Human Capital Realized in the Next Generation?* International Food Policy Research Institute. https://www.ifpri.org/publication/are-returns-mothers-human-capital-realized-next-generation

Benton, L., Mavrikis, M., Vasalou, A., Joye, N., Sumner, E., Herbert, E., ... & Raftopoulou, C. (2021). Designing for "challenge" in a large-scale adaptive literacy game for primary school children. *British Journal of Educational Technology*, *52*(5), 1862-1880. https://doi.org/1.1111/bjet.13146

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. *Science*, *350*(6257), 196–198. https://doi.org/10.1126/science.aac7427

Bettinger, E., Fairlie, R. W., Kapuza, A., Kardanova, E., Loyalka, P., & Zakharov, A. (2020). *Does EdTech substitute for traditional learning? Experimental estimates of the educational production function* [Working Paper No. w26967]. National Bureau of Economic Research. https://doi.org/10.3386/w26967

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative Inquiry*, 11(2), 226-248. https://doi.org/10.1177/1077800404273413

Blanca, M., Alarcón, R., Arnau, J., Bendayan, R., & Bono, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Boonk, L., Gijselaers, H. J. M., Ritzen, H., & Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educational Research Review*, *24*(1), 10–30. https://doi.org/10.1016/j.edurev.2018.02.001

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324. https://doi.org/10.2307/2334029

Bramley, T. (2007). Paired comparison methods. In: P.E. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 246–294). Qualifications and Curriculum Authority.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589-597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2013). *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE.

Braun, V., & Clarke, V. (2021). Thematic Analysis. In *Analysing Qualitative Data in Psychology* (pp. 128–147). SAGE Publications. https://doi.org/10.4135/9781446207536

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Breazeal, C., Morris, R. G., Gottwald, S., Galyean, T. A., & Wolf, M. (2016). Mobile Devices for Early Literacy Intervention and Research with Global Reach. *Proceesings of the Third (2016) ACM Conference on Learning @ Scale*, 11–20. https://doi.org/10.1145/2876034.2876046

Brombacher, A., Nordstrum, L., Davidson, M., Batchelder, K., Cummiskey, C., & King, S. (2014). *National Baseline Assessment for the 3Rs (Reading, Writing, and Arithmetic): Using EGRA, EGMA, and SSME in Tanzania: Study Report*. RTI International. https://ierc-publicfiles.s3.amazonaws.com/public/resources/National_Baseline_Assessment_for_the_3Rs_in_Tanzania_Study_Report_%28TO24%29.pdf

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology*, *22*(6), 723–742. https://doi.org/10.1037/0012-1649.22.6.723

Brownlee, J. (2020, July 23). *Train-Test Split for Evaluating Machine Learning Algorithms*. Machine Learning Mastery. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Buchanan, D. Ross., Miller, F. G., & Wallerstein, N. (2007). Ethical Issues in Community-Based Participatory Research: Balancing Rigorous Research With Community Participation in Community Intervention Studies. *Progress in Community Health Partnerships: Research, Education, and Action*, *1*(2), 153–160. https://doi.org/10.1353/cpr.2007.0006

Bulhan, H. A. (2015). Stages of colonialism in Africa: From occupation of land to occupation of being. *Journal of Social and Political Psychology, 3,* 239-256. https://doi.org/10.5964.jspp.v3i1.143

Burns, M. (2022, March 22). *It's Time to Update the Narrative on Educational Technology Research*. The Education and Development Forum (UKFIET). https://www.ukfiet.org/2022/its-time-to-update-the-narrative-on-educational-technology-research/

Byerengo, V. P., & Onyango, D. O. (2021). School-Family-Community Partnerships and Their Influence on Student Achievement in Public Secondary Schools in Ilemela Municipality, Tanzania. *Advances in Electronic Government, Digital Divide, and Regional Development*, 357–379. https://doi.org/10.4018/978-1-7998-6471-4.ch019

Byrnes, J. P., & Wasik, B. A. (2019). *Language and literacy development: What educators need to know*. The Guilford Press.

Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology, 43(*3), 280–293. https://doi.org/10.1080/ 17439884.2018.1498355

Camfield, J., Kobulsky, A., Paris, J., & Vonortas, N. (2007). *A Report Card for One Laptop Per Child - Closing the Digital Divide via ICTs and Education: Successes and Failures*. https://www.joncamfield.com/writing/Camfield_Report_Card_for_OLPC

Carrillo, P. E., Onofa, M., & Ponce, J. (2011). *Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador* [IDB Working Paper No. 78]. https://doi.org/10.2139/ssrn.1818756

Carroll, M., & Constantinou, F. (2023). *Teachers' Experiences of Teaching during the COVID-19 Pandemic - Research Report.* Cambridge University Press & Assessment.

Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet Based Math Assessment: What Can We Learn from Math Apps? *Journal of Educational Technology & Society*, *18(*2), 3–20.

Cerda, P., Varoquaux, G., & Kégl, B. (2018). Similarity encoding for learning with dirty categorical variables. *Machine Learning*, *107*(8-10), 1477–1494. https://doi.org/10.1007/s10994-018-5724-2

Chacky, F., Runge, M., Rumisha, S. F., Machafuko, P., Chaki, P., Massaga, J. J., Mohamed, A., Pothin, E., Molteni, F., Snow, R. W., Lengeler, C., & Mandike, R. (2018). Nationwide school malaria parasitaemia survey in public primary schools, the United Republic of Tanzania. *Malaria Journal*, *17*(452). https://doi.org/10.1186/s12936-018-2601-1

Chad-Friedman, E., Lee, Y., Liu, X., & Watson, M. W. (2019). The effects of visual arts pedagogies on children's intrinsic motivation, creativity, artistic skill, and realistic drawing ability. *The Journal of Creative Behavior*, *53*(4), 482-495. https://doi.org/10.1002/jocb.228

Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794. https://doi.org/10.1145/2939672.2939785

Cheung, S. K., Cheng, W. Y., Cheung, R. Y. M., Lau, E. Y. H., & Chung, K. K. H. (2022). Home learning activities and parental autonomy support as predictors of pre-academic skills: The mediating role of young children's school liking. *Learning and Individual Differences*, *94*, 102127. https://doi.org/10.1016/j.lindif.2022.102127

Chodura, S., Kuhn, J. T., & Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. *Zeitschrift fur Psychologie*, *223*, 129–144. https://doi.org/10.1027/2151-2604/a000211

Chuang, R., Burnett, N., & Robinson, E. (2021). *Cost-Effectiveness and EdTech: Considerations and case studies*. EdTech Hub. https://doi.org/10.5281/zenodo.5651978

Clarke, E. (2021). Methodological Pragmatism—Freedom from the Squeeze? *International Journal of Multiple Research Approaches*, *13*(3), 267–282. https://doi.org/10.29034/ijmra.v13n3a3

Clarke, V., & Braun, V. (2014). Thematic Analysis. In A. C. Michalos (Ed.), *Encyclopaedia of Quality of Life and Well-Being Research* (pp. 6626–6628). Springer. https://doi.org/10.1108/RR-06-2015-0143

Coflan, C., & Kaye, T. (2020). *Using education technology to support learners with special educational needs and disabilities in low-and middle-income countries* (Helpdesk Response No.4). EdTech Hub. https://doi.org/10.5281/zenodo.3744581

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Conn, K. (2017). Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Impact Evaluations. *Review of Educational Research, 87*(5), 863-898. https://doi.org/10.3102/0034654317712025

Conto, C. A., Akseer, S., Dreesen, T., Kamei, A., Mizunoya, S., & Rigole, A. (2020). *COVID-19: Effects of School Closures on Foundational Skills and Promising Practices for Monitoring and Mitigating Learning Loss* [Innocenti Working Paper]. UNICEF Office of Research.

Corr, M., McSharry, J., & Murtagh, E. M. (2018). Adolescent Girls' Perceptions of Physical Activity: A Systematic Review of Qualitative Studies. *American Journal of Health Promotion*, *33*(5), 806–819. https://doi.org/10.1177/0890117118818747

Cortés-Albornoz, M. C., Ramírez-Guerrero, S., García-Guáqueta, D. P., Vélez-Van-Meerbeke, A., & Talero-Gutiérrez, C. (2023). Effects of remote learning during COVID-19 lockdown on children's learning abilities and school performance: A systematic review. *International Journal of Educational Development*, *101*, 102835. https://doi.org/10.1016/j.ijedudev.2023.102835

Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education*, *43*(1), 75-96. https://doi.org/10.1080/15391523.2010.10782562

Cristia, J., Ibarraran, P., Cueto, S., Santiago, A., & Severin, E. (2012). *Technology and Child Development: Evidence from the One Laptop Per Child Program* [Discussion Paper No. 6401]. https://doi.org/10.2139/ssrn.2032444

Crompton, H., Burke, D., Jordan, K., & Wilson, S. W. G. (2021a). Learning with Technology during emergencies: A Systematic Review of K-12 Education. *British Journal of Educational Technology*, *52*(4). https://doi.org/10.1111/bjet.13114

Crompton, H., Burke, D., Jordan, K., Wilson, S., Nicolai, S., & Myers, C. (2021b). *EdTech and Emergency Remote Learning: A Systematic Review.* EdTech Hub. https://doi.org/10.5281/zenodo.4917221

Crompton, H., Chigona, A., Jordan, K., & Myers, C. (2021c). *Inequalities in Girls' Learning Opportunities via EdTech: Addressing the Challenge of Covid-19* [Working Paper 31]. EdTech Hub. https://doi.org/10.5281/ZENODO.4917252

Cuthbert, P. F. (2005). The student learning process: Learning styles or learning approaches? *Teaching in Higher Education*, *10*(2), 235–249. https://doi.org/10.1080/1356251042000337972

Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, *50*(1), 101–113. https://doi.org/10.1111/bjet.12595

Dawson, P., & Dawson, S. L. (2018). Sharing successes and hiding failures: 'reporting bias' in learning and teaching research. *Studies in Higher Education*, *43*(8), 1405-1416. https://doi.org/10.1080/03075079.2016.1258052

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384-397. https://doi.org/10.1002/sce.20194

de Melo, G., Machado, A., & Miranda, A. (2014). *The Impact of a One Laptop Per Child Program on Learning: Evidence from Uruguay* [Discussion Paper No. 8489]. https://doi.org/10.2139/ssrn.2505351

de Oliveira, J., Camacho, M., & Gisbert, M. (2014). Exploring student and teacher perception of E-textbooks in a Primary School. *Comunicar*, *21*(42), 87–95. https://doi.org/10.3916/c42-2014-08

Department for Education. (2019). *Home learning environment early years apps: parent guidance*. GOV.UK. https://www.gov.uk/government/publications/early-years-apps-pilot-home-learning-environment/home-learning-environment-early-years-apps-parent-guidance

DFID. (2016, June). *Girls' Education Challenge (GEC) Extension- Business Case Summary 30075859*. https://devtracker.dfid.gov.uk/projects/GB-1-204766/documents

DFID. (2018). *Steps to Success: Learning from the Girls' Education Challenge 2012-2017*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733765/Steps-to-Success.pdf

DFID. (2020, February). *Business Case and Summary Amendments 204766.* https://devtracker.dfid.gov.uk/projects/GB-1-204766/documents

Diendorfer, T., Seidl, L., Mitic, M., Mittmann, G., Woodcock, K., & Schrank, B. (2021). Determinants of social connectedness in children and early adolescents with mental disorder: A systematic literature review. *Developmental Review, 60*, 100960. https://doi.org/10.1016/j.dr.2021.100960

Dryden-Peterson, S., Dahya, N., & Adelman, E. (2017). Pathways to Educational Success Among Refugees: Connecting Locally and Globally Situated Resources. *American Educational Research Journal*, *54*(6), 1011–1047. https://doi.org/10.3102/0002831217714321

Duby, Z., Jonas, K., Bunce, B., Bergh, K., Maruping, K., Fowler, C., Reddy, T., Govindasamy, D., & Mathews, C. (2022). Navigating Education in the Context of COVID-19 Lockdowns and School Closures: Challenges and Resilience Among Adolescent Girls and Young Women in South Africa. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.856610

EdTech Hub. (2022, August 9). *Data for Decisions – Can technology be used to improve data collection, analysis, and planning to improve learning outcomes?* https://edtechhub.org/2022/08/09/what-we-are-learning-what-we-are-reading-data-for-decisions-can-technology-be-used-to-improve-data-collection-analysis-and-planning-to-improve-learning-outcomes/

EdTech Hub. (2023). *Using Technology to Improve Education for Marginalised Girls: Lessons in implementation from the Girls' Education Challenge* [Anonymised preprint]. https://doi.org/10.53832/edtechhub.0172

Education Development Center. (2020). *Zambia*. https://www.edc.org/sites/default/files/uploads/Zambia-QUESTT-IRI.pdf

Education Endowment Foundation (EEF). (2023, February). *Teaching and Learning Early Years Toolkit Guide*. https://educationendowmentfoundation.org.uk/education-evidence/using-the-toolkits

Edwards, S., Henderson, M., Gronn, D., Scott, A., & Mirkhil, M. (2016). Digital disconnect or digital difference? A socio-ecological perspective on young children's technology use in the home and the early childhood centre. *Technology, Pedagogy and Education*, *26*(1), 1–17. https://doi.org/10.1080/1475939x.2016.1152291

Engelmann, G., Smith, G., & Goulding, J. (2018). The Unbanked and Poverty: Predicting area-level socio-economic vulnerability from M-Money transactions. *2018 IEEE International Conference on Big Data (Big Data)*, 1357–1366. https://doi.org/10.1109/bigdata.2018.8622268

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118*(17), e2022376118. https://doi.org/10.31235/osf.io/ve4z7

Enuma. (2021). *Sekolah Enuma in Lampung and Medan: An Effective Digital-based Learning Solution for Young Children in Indonesia*. Enuma School. https://enumaschool.com/Sekolah_Enuma_Lampung_Medan_Case_EN.pdf

Eriksson, M., Ghazinour, M., & Hammarström, A. (2018). Different uses of Bronfenbrenner's ecological theory in public mental health research: what is their value for guiding public mental health policy and practice? *Social Theory & Health*, *16*(4), 414–433. https://doi.org/10.1057/s41285-018-0065-6

Erulkar, A. S., & Muthengi, E. (2009). Evaluation of Berhane Hewan: A Program to Delay Child Marriage in Rural Ethiopia. *International Perspectives on Sexual and Reproductive Health*, *35*(01), 6–14. https://doi.org/10.1363/3500609

Evans, D. K., & Popova, A. (2016). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. *The World Bank Research Observer*, *31*(2), 242–270. https://doi.org/10.1093/wbro/lkw004

Evans, D. K., & Yuan, F. (2021). What We Learn about Girls' Education from Interventions That Do Not Focus on Girls. *The World Bank Economic Review*, *36*(1), 244–267. https://doi.org/10.1093/wber/lhab007

Fàbregues, S., Sáinz, M., Romano, M. J., Escalante-Barrios, E. L., Younas, A., & López-Pérez, B.-S. (2023). Use of mixed methods research in intervention studies to increase young people's interest in STEM: A systematic methodological review. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.956300

Fadhli, M., Brick, B., Setyosari, P., Ulfa, S., & Kuswandi, D. (2020). A Meta-Analysis of Selected Studies on the Effectiveness of Gamification Method for Children. *International Journal of Instruction*, *13*(1), 845–854. https://doi.org/10.29333/iji.2020.13154a

Fairlie, R., & Loyalka, P. (2020). Schooling and COVID-19: Lessons from recent research on EdTech. *NPJ Science of Learning*, *5*(1). https://doi.org/10.1038/s41539-020-00072-6

Feeley, N., Cossette, S., Côté, J., Héon, M., Stremler, R., Martorella, G., & Purden, M. (2009). The importance of piloting an RCT intervention. *Canadian Journal of Nursing Research, 41*(2), 84–99.

Ferre, C. (2009). *Age At First Child: Does Education Delay Fertility Timing? The Case Of Kenya* [Policy Research Working Paper 4833]. World Bank. https://doi.org/10.1596/1813-9450-4833

Fietzer, A. W., & Chin, S. (2017). The Impact of Digital Media on Executive Planning and Performance in Children, Adolescents, and Emerging Adults. *Cognitive Development in Digital Contexts*, 167–180. https://doi.org/10.1016/b978-0-12-809481-5.00008-0

Fisher, D., & Frey, N. (2021). *Better learning through structured teaching: A framework for the gradual release of responsibility*. ASCD.

Flannagan, J. S., & Rockenbaugh, L. (2010). Teaching inquisitive young children how to ask good questions. *Science and Children*, *48*(4), 28-31.

Fleer, M. (2015). Pedagogical positioning in play–teachers being inside and outside of children's imaginary play. *Early Child Development and Care, 185*(11-12), 1801-1814. https://doi.org/10.1080/03004430.2015.1028393

Fomba, B. K., Talla, D. F., & Ningaye, P. (2022). Institutional Quality and Education Quality in Developing Countries: Effects and Transmission Channels. *Journal of the Knowledge Economy, 14*, 86–115. https://doi.org/10.1007/s13132-021-00869-9

Forbes. (2021, August 19). *Education apps kept kids learning for over a year, now these apps will shape our future.* https://www.forbes.com/sites/eladnatanson/2021/08/19/education-apps-kept-kids-learning-for-over-a-year--now-these-apps-will-shape-our-future/?sh=3690724f4d35

Foster, C. (2024). Methodological pragmatism in educational research: from qualitative-quantitative to exploratory-confirmatory distinctions. *International Journal of Research & Method in Education*, *47*(1), 4–19. https://doi.org/10.1080/1743727x.2023.2210063

French, B., Outhwaite, L. A., Langley-Evans, S. C., & Pitchford, N. J. (2020). Nutrition, growth, and other factors associated with early cognitive and motor development in Sub-Saharan Africa: a scoping review. *Journal of Human Nutrition and Dietetics*, *33*(5), 644–669. https://doi.org/10.1111/jhn.12795

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Fute, A., Wan, X., Oubibi, M., & Bulugu, J. B. (2022). Adult Literacy Education and Reduction of Poverty in Tanzania: A Review of Policies and their Implementation. *Journal of Education, 203*(4), 931-938. https://doi.org/10.1177/00220574221075204

Gakidou, E., Cowling, K., Lozano, R., & Murray, C. J. (2010). Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. *The Lancet*, *376*(9745), 959–974. https://doi.org/10.1016/s0140-6736(10)61257-3

Ganimian, A. J., & Murnane, R. J. (2016). Improving Education in Developing Countries: Lessons from Rigorous Impact Evaluations. *Review of Educational Research*, *86*(3), 719–755. https://doi.org/10.3102/0034654315627499

Garcia, J. (2019). *Technology and Language Learning: Assessing the influence of prior iPad experience* (Publication No. 22620299)[MA Thesis, The University of Arizona]. Proquest Dissertations Publishing. https://www.proquest.com/openview/e71ef7139f42082614396586d42cab1a/1?pq-origsite=gscholar&cbl=18750&diss=y

Garcia, M., & Fares, J. (2008). *Directions in Human Development: Youth in Africa's Labor Market*. World Bank. https://documents1.worldbank.org/curated/zh/687001468007775569/pdf/454880PUB0Box311OFFICIA0L0USE0ONLY1.pdf

Garris, W. R., Lester, L., Doran, E., Lowery, A., & Weber, A. (2018). iCollaborate or Not: Does Technology Impede Collaborative Learning among Primary Grade Students? *International Journal of Learning, Teaching and Educational Research*, *17*(5), 64-81. https://doi.org/10.26803.ijilter.17.5.5

Ghuman, S., & Lloyd, C. (2010). Teacher Absence as a Factor in Gender Inequalities in Access to Primary Schooling in Rural Pakistan. *Comparative Education Review*, *54*(4), 539–554. https://doi.org/10.1086/654832

Gimenez-Nadal, J. I., & Molina, J. A. (2012). Parents' education as a determinant of educational childcare time. *Journal of Population Economics*, *26*(2), 719–749. https://doi.org/10.1007/s00148-012-0443-7

Giraldo, J. P., Tungatarova, A., & Cooper, R. (2021, December 10). *Digital Learning Solutions | Digital Learning For Every Child*. EdTech Hub. https://edtechhub.org/2021/12/10/digital-learning-solutions-in-every-young-persons-hands/

Glewwe, P., & Miguel, E. A. (2007). The Impact of Child Health and Nutrition on Education in Less Developed Countries. In T. P. Schultz & J. Strauss (Eds.), *Handbook of Development Economics (Vol.4)*. Elsevier B.V.

Global Humanitarian Overview. (2022). *School Closures Have Severely Disrupted Education, and Remote Learning Remains Out of Reach for Many*. https://2022.gho.unocha.org/trends/school-closures-have-severely-disrupted-education-and-remote-learning-remains-out-reach-many/

Global Partnership for Education. (2023). *How civil society is helping transform education in Tanzania*. https://www.globalpartnership.org/blog/how-civil-society-helping-transform-education-tanzania#:~:text=Working%20in%20partnership%20to%20transform

Gochyyev, P., Beglar, D., Kyllonen, P., & Evans, J. (2019). *Development and validation of the Early Grade Reading and Mathematics Assessments (EGRA and EGMA)*. Educational Assessment, 24(4), 257-279. https://doi.org/10.1080/10627197.2019.1681708

Gor, B. (2023, December 7). *Implementing Edtech at Home in India: Learnings from Mindspark Usage During COVID-19*. EI Study. https://ei.study/implementing-edtech-at-home-in-india-learnings-from-mindspark-usage-during-covid-19/

Gottwald, S., Morris, R., Wolf, M., & Galyean, T. (2017). Bringing the Bottom Billion into Basic Literacy: How We Can and Why We Must. *New Directions for Child and Adolescent Development*, *2017*(158), 93–104. https://doi.org/10.1002/cad.20225

Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour*, *5*(5), 1–9. https://doi.org/10.1038/s41562-021-01212-7

GOV.UK. (2013). *Guidance - Girls' Education Challenge*. https://www.gov.uk/guidance/girls-education-challenge

Grépin, K. A., & Bharadwaj, P. (2015). Maternal education and child mortality in Zimbabwe. *Journal of Health Economics*, *44*, 97–117. https://doi.org/10.1016/j.jhealeco.2015.08.003

Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology*, *106*(1), 58-68. https://doi.org//10.1037/a0033208

Gro, J. S. (2017). *The state of the field & future direction.* Center for Curriculum Redesign. https://www.curriculumredesign.org

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). DIVERSITY: Culture, Gender, and Math. *Science*, *320*(5880), 1164–1165. https://doi.org/10.1126/science.1154094

Gulati, S. (2008). Technology-enhanced learning in developing nations: A review. *International Review of Research in Open and Distance Learning, 9(*1), 1–16. https://doi.org/10.19173/irrodl.v9i1.477

Gulliford, A., & Miller, A. (2023). Coping with life by coping with school? School non-attendance in young people. In A. Gulliford & S. Birch (Eds.), *Educational Psychology (3rd Edition)* (pp. 230–246). Routledge. https://doi.org/10.4324/9780429322815

Gulliford, A., Walton, J., Allison, K., & Pitchford, N. (2021). A qualitative investigation of implementation of app-based maths instruction for young learners. *Educational and Child Psychology*, *38*(3), 90–108. https://doi.org/10.53841/bpsecp.2021.38.3.90

Gurgand, L., Peyre, H., Écalle, J., Fischer, J., & Ramus, F. (2023). *Sibling effects on Numeracy and Literacy achievement: Evidence from Two Large French Cohorts*. https://doi.org/10.31234/osf.io/qp2az

Habyarimana, J. P., & Sabarwal, S. (2018). *Re-Kindling Learning: EReaders in Lagos* [Policy Research Working Paper No. 8665]. World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/659331544105347027/re-kindling-learning-ereaders-in-lagos

Hammond, M. (2019). What is an ecological approach and how can it assist in understanding ICT take-up? *British Journal of Educational Technology*, *51*(3), 853–866. https://doi.org/10.1111/bjet.12889

Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and techniques*. Morgan Kaufmann.

Hanushek, E. A., & Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. *Handbook of the Economics of Education*, 89–200. https://doi.org/10.2139/ssrn.1603374

Hassan, H., Islam, A., Siddique, A., & Wang, L. C. (2021). *Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh* [Discussion Paper Series No. 16525]. Institute of Labor Economics. https://doi.org/10.31235/osf.io/mhyq5

Haßler, B., Hennessy, S., Cross, A., Chileshe, E., & Machiko, B. (2014). School-based professional development in a developing context: lessons learnt from a case study in Zambia. *Professional Development in Education*, *41*(5), 806–825. https://doi.org/10.1080/19415257.2014.938355

Heath, R., & Jayachandran, S. (2017). The causes and consequences of increased female education and labor force participation in developing countries. In *The Oxford Handbook of Women and the Economy* (pp. 345–367). Oxford University Press.

Hennessy, S., Jordan, K., & Wagner, D. (2021). *Problem Analysis and Focus of EdTech Hub's Work Technology in Education in Low-and Middle-Income Countries* [Working Paper 7]. EdTech Hub. https://doi.org/10.5281/zenodo.4332693

Hennessy, S., & London, L. (2013). *Learning from International Experiences with Interactive Whiteboards* [Working Paper No. 89]. OECD Education. https://doi.org/10.1787/5k49chbsnmls-en

Herodotou, C. (2018). Young children and tablets: A systematic review of effects on learning and development. *Journal of Computer Assisted Learning*, *34*(1), 1-9. https://doi.org/10.1111/jcal.12220

Hesse-Biber, S. (2015). Mixed Methods Research. *Qualitative Health Research*, *25*(6), 775–788. https://doi.org/10.1177/1049732315580558

Hew, K. F., & Cheung, W. S. (2013). Use of Web 2.0 technologies in K-12 and higher education: The search for evidence-based practice. *Educational Research Review*, *9*, 47–64. https://doi.org/10.1016/j.edurev.2012.08.001

Higgins, S., Xiao, Z., & Katsipataki, M. (n.d.). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Pu%20332%20blications/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps. *Psychological Science in the Public Interest*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721

Holec, H. (2001). *Autonomy and foreign language learning*. Person Education.

Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1). https://doi.org/10.1140/epjds/s13688-020-00247-6

Hossain, M. M., Abdulla, F., Hai, A., Ferdous, T., Rahman, A., & Rahman, A. S. (2023). Exploring the Prevalence, Duration and Determinants of Participation in Household Chores Among Children Aged 5–17 Years in Bangladesh. *Child Indicators Research*, *16*(5), 2107–2124. https://doi.org/10.1007/s12187-023-10051-z

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01-11. https://doi.org/10.5121/ijdkp.2015.5201

Houston, S. (2014). Meta-theoretical paradigms underpinning risk in child welfare: Towards a position of methodological pragmatism. *Children and Youth Services Review*, *47*(1), 55–60. https://doi.org/10.1016/j.childyouth.2013.12.003

Howitt, D. (2016). *Introduction to Qualitative Research Methods in Psychology* (3rd Ed.). Pearson.

Hsin, C. T., Li, M. C., & Tsai, C. C. (2014). The influence of young children's use of technology on their learning: A review. *Journal of Educational Technology & Society*, *17*(4), 85-99.

Hubber, P. J., Outhwaite, L. A., Chigeda, A., McGrath, S., Hodgen, J., & Pitchford, N. J. (2016). Should Touch Screen Tablets Be Used to Improve Educational Outcomes in Primary School Children in Developing Countries? *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00839

Humanitarian Data Exchange. (2020). *United Republic of Tanzania - Humanitarian Data Exchange*. Data.humdata.org. https://data.humdata.org/group/tza

Huntington, B., Goulding, J., & Pitchford, N. J. (2023a). Expert perspectives on how educational technology may support autonomous learning for remote out-of-school children in low-income contexts. *International Journal of Educational Research Open*, *5*, 100263. https://doi.org/10.1016/j.ijedro.2023.100263

Huntington, B., Goulding, J., & Pitchford, N. (2023b). Pedagogical features of interactive apps for effective learning of foundational skills. *British Journal of Educational Technology*, *54*(5), 1273–1291. https://doi.org/10.1111/bjet.13317

International Center for Research on Women (ICRW). (2016). *A second look at the role education plays in women's empowerment.* https://www.icrw.org/wp-content/uploads/2016/10/A-Second-Look-at-the-Role-Education-Plays-in-Womens-Empowerment.pdf

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* [Working Paper No.43]. EdTech Hub. https://doi.org/10.53832/edtechhub.0070

Islam, M. S., & Grönlund. (2016). An international literature review of 1:1 computing in schools. *Journal of Educational Change*, *17(*2), 191–222. https://doi.org/10.1007/s10833-016-9271-y.

Ito, H., Keiko, K., & Makiko, N. (2021). Does Computer-aided Instruction Improve Children's Cognitive and Non-cognitive Skills?: Evidence from Cambodia. *Asian Development Review, 38*(1), 98-118. https://doi.org/10.1162/adev_a_00159

Jafari, M., & Ansari-Pour, N. (2019). Why, When and How to Adjust Your P Values? *Cell Journal (Yakhteh)*, *20*(4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Jeynes, W. (2012). A Meta-Analysis of the Efficacy of Different Types of Parental Involvement Programs for Urban Students. *Urban Education*, *47*(4), 706–742. https://doi.org/10.1177/0042085912445643

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014.

Jonassen, D. H., & Grabowski, B. L. (2012). *Handbook of Individual Differences, Learning, and Instruction*. Routledge.

Jones, A., Scanlon, E., Gaved, M., Blake, C., Collins, T., Clough, G., Kerawalla, L., Littleton, K., Mulholland, P., Petrou, M., & Twiner, A. (2013). Challenges in personalisation: supporting mobile science inquiry learning across contexts. *Research and Practice in Technology Enhanced Learning*, *8*(1), 21–42.

Jordan, K., Proctor, J., Koomar, S., & Bapna, A. (2021). *A Country-level Research Review: EdTech in Tanzania* [Working Paper]. EdTech Hub. https://docs.edtechhub.org/lib/RQGJRKVS/download/UQYUKTMB/Jordan%20et%20al_2021_A%20Country-Level%20Research%20Review.pdf

Joseph, K., & Irhene, L. (2021). The Effect of COVID-19 pandemic on the implementation of free education policy in Tanzania: A case of public secondary schools in Arusha City. *International Journal of Educational Policy Research and Review*, *8*(5). https://doi.org/10.15739/IJEPRR.21.022

Kaguo, F. E. (2011). *Factors influencing academic performance of students in community and government built secondary schools in Mbeya municipality, Tanzania* [Doctoral dissertation, Sokoine University of Agriculture]. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f509ec89dde85e988956943af482b2fbf8662385

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954-2965. https://doi.org/10.1111/jan.13031

Kalton, G. & Schuman, H. (1982) The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society Series A (General), 145,* 42–73. https://doi.org/10.2307/2981421

Kanders, K., Hickman, M., & Bazalgette, L. (2022, June). *Could toddler tech help to get more children school ready?* Nesta. https://www.nesta.org.uk/project/mapping-parenting-technology/could-toddler-tech-help-to-get-more-children-school-ready/

Karpicke, J. D. (2017b). Retrieval-Based Learning: A Decade of Progress. In J. H. Byrne (Ed.), Learning and Memory: A Comprehensive Reference (2nd ed., Volume 2) (pp. 487–514). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Kaye, T., Groeneveld, C., & Bashir, A. (2020). *Monitoring Distance Education: A Brief to Support Decision-Making in Bangladesh and Other Low- and Lower-Middle Income*

Countries (Helpdesk Response No. 30). EdTech Hub. https://doi.org/10.5281/zenodo.4140104

Keengwe, J., & Bhargava, M. (2013). Mobile learning and integration of mobile technologies in education. *Education and Information Technologies, 19(*4), 737–746. https://doi.org/10.1007/s10639-013-9250-3

Kickmeier-Rust, M. D., Hillemann, Eva-C., & Albert, D. (2014). Gamification and Smart Feedback. *International Journal of Game-Based Learning*, *4*(3), 35–46. https://doi.org/10.4018/ijgbl.2014070104

Kim, J., Gilbert, J., Yu, Q., & Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open*, *7*(1), 1-19. https://doi.org/10.1177/23328584211004183

King, S., Presley, Pouzvara, & Gove. (2019). *GLXP Data Summary | SharEd*. RTI International. https://shared.rti.org/content/global-learning-xprize-data-summary

Kishore, D., & Shah, D. (2019). *Using technology to facilitate educational attainment: Reviewing the past and looking to the future* [Background Paper Series No.23]. Pathways for Prosperity Commission.

Knoester, M., & Plikuhn, M. (2016). Influence of siblings on out-of-school reading practices. *Journal of Research in Reading*, *39*(4), 469–485. https://doi.org/10.1111/1467-9817.12059

Kolak, J., Norgate, S. H., Monaghan, P., & Taylor, G. (2021). Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. *Journal of Children and Media*, *15*(3), 410-430. https://doi.org/10.1080/17482798.2020.1844776

Korin, A. (2021). *Using EdTech to Support Learning Remotely in the Early Years: Rapid Literature Review of Evidence from the Global Response to Covid-19* [Helpdesk Response No.31]. EdTech Hub. https://doi.org/10.5281/zenodo.4746391.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kratochwill, T. R., & Levin, J. R. (Eds.). (2014). *Single-case intervention research: Methodological and statistical advances.* American Psychological Association. https://doi.org/10.1037/14376-000

Krönke, M. (2020). *Africa's digital divide and the promise of e-learning* [Policy Paper No.66]. Afrobarometer. https://www.afrobarometer.org/wp-content/uploads/migrated/files/publications/Policy%20papers/pp66-africas_digital_divide_and_the_promise_of_e-learning-afrobarometer_policy_paper-14june20.pdf

Ku, O., Chen, S. Y., Wu, D. H., Lao, A. C. C., & Chan, T-W. (2014). The Effects of Game-Based Learning on Mathematical Confidence and Performance: High Ability vs. Low Ability. *Journal of Educational Technology & Society*, *17*(3), 65–78. https://www.jstor.org/stable/10.2307/jeductechsoci.17.3.65

Kucirkova, N. (2014). iPads in early education: Separating assumptions and evidence. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00715

Kukulska-Hulme, A., Ram Ashish Giri, Saraswati Dawadi, Kamal Raj Devkota, & Gaved, M. (2023). Languages and technologies in education at school and outside of school: Perspectives from young people in low-resource countries in Africa and Asia. *Frontiers in Communication*, *8*. https://doi.org/10.3389/fcomm.2023.1081155

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., & Rozelle, S. (2013). Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. *Journal of Development Effectiveness*, *5*(2), 208–231. https://doi.org/10.1080/19439342.2013.780089

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863). https://doi.org/10.3389/fpsyg.2013.00863

Lamtey, G. (2024, January 29). *Use of smartphones still low in Tanzania*. The Citizen. https://www.thecitizen.co.tz/tanzania/news/business/use-of-smart-phones-still-low-in-tanzania-4507492

Lan, Y.-J. (2018). Technology enhanced learner ownership and learner autonomy through creation. *Educational Technology Research and Development*, *66*(4), 859–862. https://doi.org/10.1007/s11423-018-9608-8

Larkin, K. (2015). The Search for Fidelity in Geometry Apps: An Exercise in Futility? *Annual Meeting of the Mathematics Education Research Group of Australasia (MERGA)*.

Lavelle-Hill, R. (2020). *Big data psychology* [PhD Dissertation, University of Nottingham]. ResearchGate. https://www.researchgate.net/publication/349074103_Big_data_psychology

Lavelle-Hill, R., Smith, G., Mazumder, A., Landman, T., & Goulding, J. (2021). Machine learning methods for "wicked" problems: exploring the complex drivers of modern slavery. *Humanities and Social Sciences Communications*, *8*(1). https://doi.org/10.1057/s41599-021-00938-z

Lee, C. Y., & Cherner, T. S. (2015). A comprehensive evaluation rubric for assessing instructional apps. *Journal of Information Technology Education*, *14*, 21-53. http://www.jite.org/documents/Vol14/JITEV14ResearchP021-053Yuan0700.pdf

Lee, H. M., & Loo, P. A. (2021). Gamification of Learning in Early Age Education. *Journal La Edusci*, *2*(2), 44-50. https://doi.org/10.37899/journallaedusci.v2i2.380

Leigh-Hunt, N., Bagguley, D., Bash, K., Turner, V., Turnbull, S., Valtorta, N., & Caan, W. (2017). An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public Health*, *152*, 157–171. https://doi.org/10.1016/j.puhe.2017.07.035

Lemon, A., & Battersby-lennard, J. (2009). Overcoming The Apartheid Legacy In Cape Town Schools. *Geographical Review*, *99*(4), 517–538. https://doi.org/10.1111/j.1931-0846.2009.tb00445.x

Levitt, H. M. (2018). How to conduct a qualitative meta-analysis: Tailoring methods to enhance methodological integrity. *Psychotherapy Research*, *28*(3), 367-378. https://doi.org/10.1080/10503307.2018.1447708

Lewin, K. M., & Sabates, R. (2012). Who gets what? Is improved access to basic education pro-poor in Sub-Saharan Africa? *International Journal of Educational Development, 32*(4), 517–528. https://doi.org/10.1016/j.ijedudev.2012.02.013

Lewis, I. M. (2021). *Witchcraft - Witchcraft in Africa and the world*. Encyclopaedia Britannica. https://www.britannica.com/topic/witchcraft/Witchcraft-in-Africa-and-the-world

Lindskog, A. (2011). *The Effect of Older Siblings' Literacy on School Entry and Primary School Progress in the Ethiopian Highlands.* RePEc: Research Papers in Economics.

Lippard, C. N., La Paro, K. M., Rouse, H. L., & Crosby, D. A. (2017). A Closer Look at Teacher–Child Relationships and Classroom Emotional Context in Preschool. *Child & Youth Care Forum*, *47*(1), 1–21. https://doi.org/10.1007/s10566-017-9414-1

Lloyd, C., Mete, C., & Grant, M. (2007). Rural girls in Pakistan: Constraints of policy and culture. In M. A. Lewis & M. E. Lockhead (Eds.), *Exclusion, Gender and Education: Case Studies from the Developing World.* Centre for Global Development.

Long, H., & Pang, W. (2016). Family socioeconomic status, parental expectations, and adolescents' academic achievements: a case of China. *Educational Research and Evaluation*, *22*(5-6), 283–304. https://doi.org/10.1080/13803611.2016.1237369

Longley, S. (2021). *Scaling personalised learning technology in Malawi*. EdTech Hub. https://edtechhub.org/sandboxes/scaling-personalised-learning-technology-in-malawi/

Longley, S. (2022). *Technology to Advance Data Use & Decision-Making in Education*. EdTech Hub. https://edtechhub.org/our-topic-areas/data-for-decisions/

Lopez, M., Ruiz, M. O., Rovnaghi, C. R., Tam, G. K-Y., Hiscox, J., Gotlib, I. H., Barr, D. A., Carrion, V. G., & Anand, K. J. S. (2021). The social ecology of childhood and early life adversity. *Pediatric Research*, *89*(2). https://doi.org/10.1038/s41390-020-01264-x

Lucas, K. (2011). Making the connections between transport disadvantage and the social exclusion of low income populations in the Tshwane Region of South Africa. *Journal of Transport Geography*, *19*(6), 1320–1334. https://doi.org/10.1016/j.jtrangeo.2011.02.007

Lundberg, S. M., & Lee, S-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1705.07874

Lurvink, A.-F., & Pitchford, N. (2023). Introduction of an EdTech intervention to support learning of foundational skills in Sierra Leone: Policy, teacher, and community perspectives. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1069857

Lyanga, A. A., & Chen, M.-K. (2020). The Impacts of Fee-Free Education Policy in Junior Secondary Schools in Tanzania. *Asian Journal of Education and Social Studies*, *13*(3), 36–47. https://doi.org/10.9734/ajess/2020/v13i330333

Lynch, P., Singal, N., & Francis, G. A. (2024). Educational technology for learners with disabilities in primary school settings in low- and middle-income countries: A systematic literature review. *Educational Review*, *76*(2), 405-431. https://doi.org/10.1080/00131911.2022.2035685

Ma, Y., Fairlie, R., Loyalka, P., & Rozelle, S. (2020). *Isolating the "Tech" from EdTech: Experimental Evidence on Computer Assisted Learning in China* [Working Paper w26953]. National Bureau of Economic Research. https://doi.org/10.3386/w26953

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, *9*(3), 3351-3354.

Major, L., & Francis, G. (2020). *Technology-supported personalised learning: Rapid evidence review.* EdTech Hub. https://edtechhub.org/wp-content/uploads/2020/09/Rapid-Evidence-Revie w_-Technology-supported-personalised-learning.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). https://doi.org/10.1111/bjet.13116

Marshall, N., Shaw, K., Hunter, J., & Jones, I. (2020). Assessment by comparative judgement: An application to secondary statistics and English in New Zealand. *New Zealand Journal of Educational Studies*, *55*(1), 49-71. https://doi.org/10.1007/s40841-020-00163-3

Mbilu, E. M. (2019). *Non-Formal Education and Quality Basic Education for All in Tanzania: A study of the Effectiveness of Non-Formal Education Programs for Children of School-going Age in Korogwe District* [MPhil Thesis, University of Oslo]. https://www.duo.uio.no/bitstream/handle/10852/73856/5/MPhil-CIE_EDU4491_Thesis.pdf

McCarthy, A. S., & Pearlman, R. (2022). Multiplying Siblings: Exploring the Trade-off Between Family Size and Child Education in Rural Bangladesh. *Journal of Development Studies*, *58*(9), 1831–1856. https://doi.org/10.1080/00220388.2022.2048652

McCormack, B., Rycroft-Malone, J., DeCorby, K., Hutchinson, A. M., Bucknall, T., Kent, B., ... & Wilson, V. (2013). A realist review of interventions and strategies to promote evidence-informed healthcare: a focus on change agency. *Implementation Science*, *8*(1), 1-12. https://doi.org/10.1186/1748-5908-8-107

McEwan, P. J. (2015). Improving Learning in Primary Schools of Developing Countries: A meta-analysis of randomised experiments. *Review of Educational Research*, *85*(3), 353–394. https://doi.org/10.3102/0034654314553127

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.

McGough, J. J., & Faraone, S. V. (2009). Estimating the Size of Treatment Effects: Moving Beyond P Values. *Psychiatry (Edgmont)*, *6*(10), 21–29.

Menashy, F., & Zakharia, Z. (2017). *Investing in the crisis: Private participation in the education of Syrian refugees.* Education International. https://www.right-to-education.org/sites/right-to-education.org/files/resource-attachments/EI_Investing%20in%20Crisis_Private%20participation%20in%20education%20of%20Syrian%20refugees_April%202017_EN.pdf

Meroka-Mutua, A. K., Mwanga, D., & Olungah, C. O. (2020). Assessing the role of law in reducing the practise of FGM/C in Kenya. *Evidence to End FGM/C: Research to Help Girls and Women Thrive*. Population Council. https://doi.org/10.31899/rh12.1026

Meyer, M., Zosh, J. M., McLaren, C., Robb, M., McCaffery, H., Golinkoff, R. M., Hirsh-Pasek, K., & Radesky, J. (2021). How educational are “educational” apps for young children? App store content analysis using the Four Pillars of Learning framework. *Journal of Children and Media*, *15*(4), 1–23. https://doi.org/10.1080/17482798.2021.1882516

Microsoft Bing. (2020, February 12). *Building Footprints - Bing Maps*. https://www.microsoft.com/en-us/maps/bing-maps/building-footprints

Miglani, N., & Burch, P. (2018). Educational Technology in India: The Field and Teacher's Sensemaking. *Contemporary Education Dialogue*, *16*(1), 26–53. https://doi.org/10.1177/0973184918803184

Miheretu, A. (2019). *Meeting the Academic and Social-Emotional Needs for Nigeria's Out-of-School Children: What works and what doesn't for an accelerated learning program* [Research Brief]. International Rescue Committee. https://inee.org/resources/meeting-academic-and-social-emotional-needs-nigerias-out-school-children

Mindspark. (2021). *VELA Microschool's EA Case Study*. https://www.mindspark.org/vela-microschool-ea

Mokoena, V. T. (2015). *Problems affecting the management of Public Adult Learning Centres at Sabie circuit in Mpumalanga Province in South Africa* [MEd Dissertation, University of Limpopo]. http://hdl.handle.net/10386/1550

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*(mar196), h1258. https://doi.org/10.1136/bmj.h1258

More, A. S., & Rana, D. P. (2017). Review of Random Forest classification techniques to resolve data imbalance. *1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 72–78. https://doi.org/10.1109/ICISIM.2017.8122151

Moss, C. (2020, Jan 22). *18 large-scale EdTech initiatives on our radar in 2020*. EdTech Hub. https://edtechhub.org/2020/01/22/18-large-scale-edtech-initiatives-on-our-radar-in-2020/

Moumen, H. (2023, July 21). *Children with Disabilities in Eastern and Southern Africa: A statistical overview of their well-being*. UNICEF Data. https://data.unicef.org/resources/children-with-disabilities-in-eastern-and-southern-africa-a-statistical-overview-of-their-well-being/#:~:text=Nearly%2029%20million%20children%20with%20disabilities%20live%20in%20Eastern%20and%20Southern%20Africa.

Mouza, C., & Cavalier, A. (2013). The Role of One-to-One Computing in the Education of at-Risk High-School Students, *Emerging Technologies for the Classroom*, 145-159. https://doi.org/10.1007/978-1-4614-4696-5_10

Muhammad, N. M., Schneider, M., Hill, A., & Yau, D. M. (2019, March). The Negative Impacts of EdTech: EQ Perspectives. In *Society for Information Technology & Teacher Education International Conference* (pp. 1066-1071). Association for the Advancement of Computing in Education (AACE).

Müller, T. (2019). *She Works Hard for the Money: Tackling Low Pay in Sectors Dominated by Women – Evidence from Health and Social Care* [Working Paper 2019.11]. ETUI. https://doi.org/10.2139/ssrn.3448991

Mundy, K., & Hares, S. (2020). *Equity-Focused Approaches to Learning Loss during COVID-19*. Center for Global Development. https://www.cgdev.org/blog/equity-focused-approaches-learning-loss-during-covid-19

Muñoz-Najar, A., Gilberto, A., Hasan, A., Cobo, C., Azavedo, J., & Akmal, M. (2021). *Remote Learning during COVID-19: Lessons from Today, Principles for Tomorrow.* World Bank.

https://documents1.worldbank.org/curated/en/160271637074230077/pdf/Remote-Learning-During-COVID-19-Lessons-from-Today-Principles-for-Tomorrow.pdf

Musomi, M., & Swadener, B. B. (2017). Enhancing Feminism and Childhoods in Kenya Through Stronger Education Policy, Access, and Action. In *Feminism(s) in Early Childhood* (pp. 75–87). Springer.

Musso, M. F., Cascallar, E. C., Bostani, N., & Crawford, M. (2020). Identifying Reliable Predictors of Educational Outcomes Through Machine-Learning Predictive Modeling. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00104

National Early Literacy Panel. (2008). *Developing Early Literacy: Report of the National Early Literacy Panel.* National Institute for Literacy.

Naylor, R., & Gorgen, K. (2020). *Overview of emerging country-level response to providing educational continuity under COVID-19: What are the lessons learned from supporting education for marginalised girls that could be relevant for EdTech responses to COVID-19 in lower- and middle-income countries?* EdTech Hub. https://doi.org/10.5281/zenodo.4706059

Nica-Avram, G., Harvey, J., Smith, G., Smith, A., & Goulding, J. (2021). Identifying food insecurity in food sharing networks via machine learning. *Journal of Business Research*, *131*, 469–484. https://doi.org/10.1016/j.jbusres.2020.09.028

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*, 102841. https://doi.org/10.1016/j.ijedudev.2023.102841

Njie, H. (2016). The Interaction of Economic Livelihood Strategies and Literacy and Numeracy Practices of Urban Gambian Women with Low Educational Attainments.

International Journal of Education and Literacy Studies, *4*(3). https://doi.org/10.7575/aiac.ijels.v.4n.3p.73

Njie, H., Manion, C., & Badjie, M. (2015). Girls' Familial Responsibilities and Schooling in The Gambia. *International Education Studies*, *8*(10). https://doi.org/10.5539/ies.v8n10p48

Ochoa, A. A. C., & Hernandez, A. S. (2023). *Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1 by complementing it with other banks of items.* UNESCO Institute for Statistics.

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education.* OECD Publishing. https://doi.org/10.1787/19963777

OECD, UNESCO, UNICEF, & World Bank. (2022). *From learning recovery to education transformation*. UNICEF. https://www.unicef.org/blog/learning-recovery-education-transformation

Ofori, F., Maina, E., & Gitonga, R. (2020). Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review. *Journal of Information and Technology*, *4*(1), 33–55.

Ogbonna, M. N. (2015). Role of internet usage in nomadic education: A strategy for achieving development in Nigeria beyond 2020. *Knowledge Review*, *33*(1), 1–5. https://www.globalacademicgroup.com/journals/knowledge%20review/Ogbonna4.pdf

Ombati, V., & Ombati, M. (2017). Gender Inequality in Education in sub-Saharan Africa. *Journal of Women's Entrepreneurship and Education*, *3-4*, 114–136.

onebillion. (2020). *COVID-19 UPDATE: How we can keep children learning*. https://onebillion.org/news/2020/03/30/covid-19-update-how-we-can-keep-children-learning/

onebillion. (2022). *onebillion: onetab – one tablet for reading and numeracy*. Onebillion.org. https://onebillion.org/onetab/

Orchid Project. (2021). *Intersection between Female Genital Cutting and Education: A discussion brief for the Global Education Summit* (pp. 1–12). https://www.orchidproject.org/wp-content/uploads/2021/07/Intersection-Between-Female-Genital-Cutting-and-Education.pdf

Orozco-Olvera, V. H., & Rascon-Ramirez, E. G. (2022). *Improving Enrollment and Learning Through Videos and Mobiles: Experimental Evidence from Northern Nigeria* [Working Paper No. 10413]. World Bank. https://doi.org/10.2139/ssrn.4221220

Osborne, J., & Hennessy, S. (2003). *Literature Review in Science Education and the Role of ICT: Promise, Problems and Future Directions*. https://telearn.hal.science/hal-00190441v1/document

Oulo, B., Sidle, A. A., Kintzi, K., Mwangi, M., & Akello, I. (2021). *Understanding the Barriers to Girls' School Return: Girls' Voices from the Frontline of the COVID-19 Pandemic in East Africa*. AMPLIFY Girls.

Outhwaite, L. A. (2019). *The use of interactive maths apps to support early mathematical development in UK and Brazilian primary school children* [PhD Thesis, University of Nottingham].

Outhwaite, L. A., Early, E., Herodotou, C., & Jo Van Herwegen. (2023). Understanding how educational maths apps can enhance learning: A content analysis and qualitative comparative analysis. *British Journal of Educational Technology*, *54*(5), 1292–1313. https://doi.org/10.1111/bjet.13339

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2019a). Raising early achievement in math with interactive apps: A randomized control trial. *Journal of Educational Psychology*, *111*(2), 284-298. https://doi.org/10.1037/edu0000286

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). Closing the gap: Efficacy of a tablet intervention to support the development of early mathematical skills in UK primary school children. *Computers & Education*, *108*, 43–58. https://doi.org/10.1016/j.compedu.2017.01.011

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2019b). A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. *International Journal of Research & Method in Education*, *43*(3), 225–242. https://doi.org/10.1080/1743727x.2019.1657081

Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2020). Language counts when learning mathematics with interactive apps. *British Journal of Educational Technology, 51*(6), 2326–2339. https://doi.org/10.1111/bjet.12912

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: what educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11(*3), 256-277. https://doi.org/10.1504/IJMLO.2017.085338

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*, *22*(6), 3147–3165. https://doi.org/10.1007/s10639-017-9579-0

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the android Google Play for Greek preschoolers: A systematic review. *Computers & Education, 116*, 139–160. https://doi.org/10.1016/j.Compedu.2017.09.007

Paradowski, M. B. (2015). Holes in SOLEs: re-examining the role of EdTech and 'minimally invasive education' in foreign language learning and teaching. *English Lingua Journal*, *1*(1), 37-60.

Park, H. M. (2003). *Understanding the statistical power of a test.* http://www.indiana.edu/*statmath/stat/all/power/power. html#One-way

Passey, D., Laferrière, T., Ahmad, Y.-A., Bhowmik, M., Gross, D., Price, J., & Shonfeld, M. (2016). Educational Digital Technologies in Developing Countries Challenge Third Party Providers. *Journal of Educational Technology & Society, 19(*3), 121–133.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review--a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, *10 Suppl 1*(1), 21–34. https://doi.org/10.1258/1355819054308530

Pegrum, M., Howitt, C., & Striepe, M. (2013). Learning to take the tablet: How pre-service teachers use iPads to facilitate their learning. *Australasian Journal of Educational Technology*, *29*(4). https://doi.org/10.14742/ajet.187

Pezzulo, C., Alegana, V. A., Christensen, A., Bakari, O., & Tatem, A. J. (2022). Understanding factors associated with attending secondary school in Tanzania using household survey data. *PLOS ONE*, *17*(2), e0263734. https://doi.org/10.1371/journal.pone.0263734

Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology*, *6*(485), 1-12. https://doi.org/10.3389/fpsyg.2015.00485

Pitchford, N. J. (2023). Customised E-Learning Platforms. In T. Madon & A.J. Gadgil (Eds.), *Introduction to Development Engineering: A Framework with Applications from the Field* (pp. 269-292). Springer. https://doi.org/10.1007/978-3-030-86065-3

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. https://doi.org/10.1111/desc.12864

Pitchford, N. J., Kamchedzera, E., Hubber, P. J., & Chigeda, A. L. (2018). Interactive Apps Promote Learning of Basic Mathematics in Children With Special Educational Needs and Disabilities. *Frontiers in Psychology*, *9*(262). https://doi.org/10.3389/fpsyg.2018.00262

Pitchford, N. J., & Outhwaite, L. A. (2016). Can Touch Screen Tablets be Used to Assess Cognitive and Motor Skills in Early Years Primary School Children? A Cross-Cultural Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01666

Pitchford, N. J., & Outhwaite, L. A. (2016). The use of tablet technology to support development of early mathematical skills: A cross cultural comparison. In N. Kucirkova and G. Falloon, (eds.), *Apps, Technology and Younger Learners: International Evidence for Teaching* (pp. 121-134). Routledge.

Pitchford, N. J., Papini, C., Outhwaite, L. A., & Gulliford, A. (2016). Fine motor skills predict maths ability better than they predict reading ability in the early primary school years. *Frontiers in Psychology*, *7*, 783. https://doi.org/10.3389/fpsyg.2016.00783

Plump, C. M., & LaRosa, J. (2017). Using Kahoot! in the classroom to create engagement and active learning: A game-based technology solution for eLearning novices. *Management Teaching Review*, *2*(2), 151-158. https://doi.org/10.1177/2379298116689783

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 281-300. https://doi.org/10.1080/0969594X.2012.665354

Porter, G. (2014). Transport Services and Their Impact on Poverty and Growth in Rural Sub-Saharan Africa: A Review of Recent Research and Future Research Needs. *Transport Reviews*, *34*(1), 25–45. https://doi.org/10.1080/01441647.2013.865148

Porter, J. (2018). Entering Aladdin’s Cave: Developing an app for children with Down syndrome. *Journal of Computer Assisted Learning*, *34*(4), 429–439. https://doi.org/10.1111/jcal.12246

Power, T., Buckler, A., Ebubedike, M., Tengenesha, M., Jama, M., Ndlovu, A., ... & Mubaira, S. (2021). *Community Help for Inclusive Learning and Development (CHILD): A Study of How Mobile Phones Were Used to Recruit and Equip Community Volunteers to Support Children's Learning During Covid-19 School Closures in Zimbabwe.* EdTech Hub. https://docs.edtechhub.org/lib/HZRHPNI5/download/VU88LKN8/Community%20Help%20for%20Inclusive%20Learning%20and%20Development_Final.pdf

Psacharopoulos, G., & Patrinos, H. A. (2002). *Returns to Investment in Education*. World Bank.

Putra, A. S., Warnars, H. L. H. S., Abbas, B. S., Trisetyarso, A., Suparta, W., & Kang, C. H. (2018, September). Gamification in the e-Learning Process for children with Attention Deficit Hyperactivity Disorder (ADHD). In *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)* (pp. 182-185). IEEE.

Putrawangsa, S., & Hasanah, U. (2018). Integration of digital technology in learning in the industrial age 4.0 [in Bahasa]. *Jurnal Tatsqif*, *16*(1), 42–54.

Qureshi, M. I., Khan, N., Gillani, S. M. A. H., & Raza, H. (2020). A Systematic Review of Past Decade of Mobile Learning: What We Learned and Where to Go. *International Journal of Interactive Mobile Technologies (IJIM)*, *14*(06), 67. https://doi.org/10.3991/ijim.v14i06.13479

Raymer, R. (2013, September 13). The rock stars of eLearning: An interview with Karl Kapp. *eLearn Magazine*. http://elearnmag.acm.org/archive.cfm?aid=2524223

Reilly, A. (2023). *Girls' Education Challenge Portfolio in Practice (4): Strengthening teaching and learning across a portfolio: A proven tool for self-assessment and action*. UK Aid.

https://girlseducationchallenge.org/media/dodb2fx2/gec_pip_4_self_assessment_tools_final.pdf

Rescher, N. (1977). *A System of Pragmatic Idealism, Volume I: Human knowledge in idealistic perspective*. New York University Press.

Rescher, N. (2017). *Pragmatism: The restoration of its scientific roots*. Routledge.

Richards, J. C. (2015). Technology in language teaching today. *Indonesian JELT*, *10*(1), 18-32. https://doi.org/10.25170/ijelt.v10i1.1506

Rising Academies. (2020, June 5). *Rising Academies - Rising On Air*. https://www.risingacademies.com/onair

Rising Academies. (2023). *Rising Academies*. https://www.risingacademies.com/

Roberts-Tyler, E. J., Roberts, S., Watkins, R., Hughes, J. C., Hastings, R. P., & Gillespie, D. (2023). Effects of implementation support on children's reading outcomes following an online early reading programme: A cluster-randomised controlled trial. *British Journal of Educational Technology*, *54*(5), 1373–1396. https://doi.org/10.1111/bjet.13312

Robson, C. (2002). *Real World Research* (2nd Ed.). Blackwell.

Rodriguez-Segura, D. (2022). EdTech in Developing Countries: A Review of the Evidence. *The World Bank Research Observer, 37*(2). https://doi.org/10.1093/wbro/lkab011

Roseland, M. (2012). *Toward sustainable communities: Solutions for citizens and their governments*. New Society Publishers.

Ryan, D. (2016). Using tablet technology for personalising learning. *Journal of Research in Special Educational Needs*, *16*(S1), 1071–1077. https://doi.org/10.1111/1471-3802.12252

Rymanowicz, K. A., Moyses, K. J., & Zoromski, K. S. (2020). School Readiness. *Encyclopedia of Infant and Early Childhood Development*, 55–64. https://doi.org/10.1016/b978-0-12-809324-5.21836-1

Sampson, R., Johnson, D., Somanchi, A., Barton, H., Ruchika, J., Seth, M., & Shotland, M. (2019). *Insights from rapid evaluations of EdTech products.* Central Square Foundation. https://www.centralsquarefoundation.org/EdTech-Lab-Report-November-2019.pdf

Samuels, F., Leon-Himmelstine, C., Marcus, R., & Myers, C. (2022). *How to Design EdTech Programmes That Lead to Gender-Transformative Change* [Position Paper]. EdTech Hub. https://docs.edtechhub.org/lib/DSG9QF4F/download/XNIJ7GJ5/Samuels%20et%20al.%20-%202022%20-%20How%20to%20Design%20EdTech%20Programmes%20%20That%20Lead%20to%20Gend.pdf

Save the Children. (2021, April 26). *393 million children unable to read: The world's shocking lost potential.* Save the Children International. https://www.savethechildren.net/news/393-million-children-unable-read-world%E2%80%99s-shocking-lost-potential

Scikit Learn. (2023). *sklearn.dummy.DummyClassifier — scikit-learn 0.23.2 documentation*. https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html

Segal, M. R. (2004). *Machine Learning Benchmarks and Random Forest Regression*. UCSF: Center for Bioinformatics and Molecular Biostatistics. https://escholarship.org/content/qt35x3v9t4/qt35x3v9t4_noSplash_3bc7fbb8348b76e0ad2a408fe58dfd94.pdf

Shah, N. (2011). Special Education Pupils Find Learning Tool in iPad Applications. *Education Week, 30*(22).

Shannon-Baker, P. (2016). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research, 10*(4), 319–334. https://doi.org/10.1177/1558689815575861

Shukia, R., Katabaro, J., & Mgumia, J. (2019). *Social Emotional Impact Assessment Draft Report* [Unpublished report]. Prepared for UNESCO, Tanzania.

Sitali, M. M. (2006). Interactive Radio Instruction. *The Fourth Pan-Commonwealth Forum on Open Learning (PCF4).*

Skene, K., O'Farrelly, C. M., Byrne, E. M., Kirby, N., Stevens, E. C., & Ramchandani, P. G. (2022). Can guidance during play enhance children's learning and development in educational contexts? A systematic review and meta-analysis. *Child Development, 93*(4), 1162-1180. https://doi.org/10.1111/cdev.13730

Smith, G., Phiri, K., Mwangi, M., Phiri, F., Goulding, J., & Pitchford, N. J. (n.d.). *Dietary diversity and soil-type indicators of primary education outcomes in Malawi* (in preparation).

Sokal, L., & Piotrowski, C. (2011). My Brother's Teacher? Siblings and Literacy Development in the Home. *Education Research International, 2011*(2), 1–6. https://doi.org/10.1155/2011/253896

Spaull, N. (2012). *Malawi at a glance; SACMEQ at a glance*. Research on Socio-Economic Policy (RESEP). https://resep.sun.ac.za/projects/sacmeq-at-a-glance/

Spaull, N., & Taylor, S. (2015). Access to What? Creating a Composite Measure of Educational Quantity and Educational Quality for 11 African Countries. *Comparative Education Review, 59*(1), 133–165. https://doi.org/10.1086/679295

Spieth, P. M., Siepmann, T., Kubasch, A. S., Penzlin, A. I., Illigens, B. M.-W., & Barlinn, K. (2016). Randomized Controlled Trials – a Matter of Design. *Neuropsychiatric Disease*

and Treatment, *12*(12), 1341–1349.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4910682/

Statista. (2021a, July 6). *Mobile education app downloads worldwide.* https://www.statista.com/statistics/1128262/mobile-education-app-downloads-worldwide-platforms-millions/

Statista (2021b, Apr 7). *Global digital population as of January 2021.* https://www.statista.com/statistics/617136/digital-population-worldwide/

Stubbé, H., Badri, A., Telford, R., Oosterbeek, S., & van der Hulst, A. (2017). Formative evaluation of a mathematics game for out-of-school children in Sudan. *Simulation and Serious Games for Education*, 61–79.

Stubbé, H., Badri, A., Telford, R., van der Hulst, A., & van Joolingen, W. (2016). E-Learning Sudan, Formal Learning for Out-of-School Children. *Electronic Journal of E-Learning, 14*(2), 136–149.

Suminar, D. R., & Wardana, N. D. (2018). Thematic Analysis of The Symbolic And Imaginary Play Indonesian Children. *The International Journal of Social Sciences and Humanities Invention*, 5(11), 5066-5071. https://doi.org/10.18535/IJSSHI/V5I11.01

Sun, J., & Shen, Z. (2022). Research on Improved Random Forest Algorithm for Highly Unbalanced Data. *Journal of Physics: Conference Series*, *2333*(1), 012007. https://doi.org/10.1088/1742-6596/2333/1/012007

Tamim, R. M., Borokhovski, E., Pickup, D., Bernard, R. M., & El Saadi, L. (2015). *Tablets for Teaching and Learning: A Systematic Review and Meta-Analysis.* http://oasis.col.org/bitstream/handle/11599/1012/2015_Tamim-et-al_Tablets-for-Teaching-andLearning.pdf

Tauson, M. & Stannard, L. (2018). *EdTech for learning in emergencies and displaced settings.* Save the Children.

https://resourcecentre.savethechildren.net/sites/default/files/documents/edtech-learning.pdf.

Tavera, G. F., & Casinillo, L. F. (2020). Knowledge Acquisition Practices and Reading Comprehension Skills of the Learners in Hilongos South District, Leyte Division, Philippines. *JPI (Jurnal Pendidikan Indonesia)*, *9*(3), 533. https://doi.org/10.23887/jpi-undiksha.v9i3.28114

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research : integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE.

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 and Busara Center for Behavioural Economics. https://doi.org/10.53832/edtechhub.0048

The Commonwealth. (2014). *Female Genital Mutilation: The Role of Education*. https://www.ungei.org/sites/default/files/2021-02/FGM-The-Role-Of-Education-2016-eng.pdf

The Economist. (2018, February 22). *A fuss over Freemasons in Africa.* https://www.economist.com/middle-east-and-africa/2018/02/22/a-fuss-over-freemasons-in-africa

Thinley, S., & Rui, Y. (2023). EdTech to support Out-of-School Children and Adolescents [Helpdesk Response 167]. *EdTech Hub.* https://doi.org/10.53832/edtechhub.0163

Tikly, L., & Barrett, A. M. (2011). Social justice, capabilities and the quality of education in low income countries. *International Journal of Educational Development*, *31*(1), 3–14. https://doi.org/10.1016/j.ijedudev.2010.06.001

Timmons, K., & Pelletier, J. (2015). Understanding the importance of parent learning in a school-based family literacy programme. *Journal of Early Childhood Literacy*, *15*(4), 510–532. https://doi.org/10.1177/1468798414552511

Ting, Y. L. (2015). Tapping into students' digital literacy and designing negotiated learning to promote learner autonomy. *The Internet and Higher Education*, *26*, 25-32. https://doi.org/10.1016/j.iheduc.2015.04.004

Tom, C. (2018). *Navajo Teachers' Experiences in Implementing Technology into the Response to Intervention Program* (Publication No. 10931526)[PhD dissertation, Capella University]. Proquest Dissertations Publishing. https://www.proquest.com/docview/2100024782

Toub, T. S., Rajan, V., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Guided play: A solution to the play versus learning dichotomy. In D. C. Geary and D. B. Berch (eds.), *Evolutionary Perspectives on Child Development and Education* (pp. 117-141). Springer. https://doi.org/10.1007/978-3-319-29986-0

Trainor, L. R., & Bundon, A. (2020). Developing the craft: reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *13*, 1-22. https://doi.org/10.1080/2159676X.2020.1840423

UK Aid. (2021). *Girls' Education Challenge*. https://girlseducationchallenge.org/countries/country/tanzania

UN Women. (2015). *Progress of the World's Women 2015-2016: Transforming Economies, Realizing Rights*. https://progress.unwomen.org/en/2015/chapter2/

UNESCO. (2015). *Gender and education for all 2000–2015: Achievements and challenges.* https://unesdoc.unesco.org/ark:/48223/pf0000234809

UNESCO. (2019). *GEM Report: Migration, displacement and education- Building bridges, not walls.* https://gem-report-2019.unesco.org/

UNESCO. (2020). *GEM Report: A new generation: 25 years of efforts for gender equality in education 2020 - Gender Report*. https://gem-report-2020.unesco.org/gender-report/

UNESCO. (2021). *Pandemic-related disruptions to schooling and impacts on learning proficiency indicators: a focus on the early grades*. https://unesdoc.unesco.org/ark:/48223/pf0000377781

UNESCO. (2022a). *Global Education Monitoring Report - Gender Report: Deepening the debate on those still left behind*. https://unesdoc.unesco.org/ark:/48223/pf0000381329

UNESCO. (2022b, October 3). *GEM Report: More boys than girls have been out of school since 2007*. https://world-education-blog.org/2022/10/03/more-boys-than-girls-have-been-out-of-school-since-2007/

UNESCO. (2022c, September). *New estimation confirms out-of-school population is growing in Sub-Saharan Africa* [Policy Paper]. https://www.unesco.org/gem-report/en/2022-out-school#:~:text=It%20is%20estimated%20that%20244,and%20125.5%20million%20were%20boys.

UNESCO. (2023, April 20). *Nearly US$100 billion finance gap for countries to reach their education targets.* https://www.unesco.org/en/articles/nearly-us100-billion-finance-gap-countries-reach-their-education-targets

UNESCO. (2024a, February 13). *GEM Report: Radio delivers education at a low cost to hard-to-reach population*. https://world-education-blog.org/2024/02/13/radio-delivers-education-at-a-low-cost-to-hard-to-reach-population/

UNESCO. (2024b). *United Republic of Tanzania (Mainland and Zanzibar): Education Country Brief.* https://www.iicba.unesco.org/en/node/111#:~:text=Within%20Tanzania%2C%20the%20National%20Examination

UNICEF. (2018). *Harmful practices*. https://www.unicef.org/protection/harmful-practices

UNICEF. (2020). *Getting Herder Families and Communities Ready For School: A lifelong learning approach* (pp. 1–10). https://www.unicef.org/mongolia/media/3961/file/Getting%20herder%20and%20families%20and%20communities%20ready%20for%20school.pdf

UNICEF. (2021a, December). *The world's nearly 240 million children living with disabilities are being denied basic rights*. UNICEF UK. https://www.unicef.org.uk/press-releases/the-worlds-nearly-240-million-children-living-with-disabilities-are-being-denied-basic-rights-unicef/

UNICEF. (2021b, November 10). *Nearly 240 million children with disabilities around the world, UNICEF's most comprehensive statistical analysis finds*. https://www.unicef.org/bulgaria/en/press-releases/nearly-240-million-children-disabilities-around-world-unicefs-most-comprehensive

UNICEF. (2022a). 70 percent of 10-year-olds in 'learning poverty' unable to read and understand a simple text. https://www.unicef.org/press-releases/70-cent-10-year-olds-learning-poverty-unable-read-and-understand-simple-text

UNICEF. (2022b). *UNICEF: Only a third of 10-year-olds globally are estimated to be able to read and understand a simple written story*. https://www.unicef.org/bulgaria/en/press-releases/unicef-only-third-10-year-olds-globally-are-estimated-be-able-read-and-understand#:~:text=Sofia%2C%2019%20September%202022%20%E2%80%93%20Only

United Nations. (2023). *Commodities at a glance: Special issue on access to energy in sub-Saharan Africa*. UNCTAD. https://unctad.org/publication/commodities-glance-special-issue-access-energy-sub-saharan-africa#:~:text=In%20sub%2DSaharan%20Africa%20alone

United Nations Development Programme (UNDP). (2023). *Goal 4: Quality Education.* https://www.undp.org/sustainable-development-goals/quality-education

USAID. (2009). *Learning at Taonga Market at Grade 2: An Evaluation of Interactive Radio Instruction in GRZ Schools in 2008*. https://pdf.usaid.gov/pdf_docs/PDACP433.pdf

Vaiopoulou, J., Papadakis, S., Sifaki, E., Kalogiannakis, M., & Stamovlasis, D. (2023). Classification and evaluation of educational apps for early childhood: Security matters. *Education and Information Technologies*, *28*, 2547–2578. https://doi.org/10.1007/s10639-022-11289-w

van Bergen, E., van Zuijen, T., Bishop, D., & de Jong, P. F. (2017). Why Are Home Literacy Environment and Children's Reading Skills Associated? What Parental Skills Reveal. *Reading Research Quarterly*, *52*(2), 147–160. https://doi.org/10.1002/rrq.160

Vanbecelaere, S., Cornillie, F., Depaepe, F., Guerrero, R. G., Mavrikis, M., Vasalou, M., & Benton, L. (2020a). Technology-mediated personalised learning for younger learners. *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*, 126–134. https://doi.org/10.1145/3397617.3398059

Vanbecelaere, S., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2021). The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, noncognitive and efficiency outcomes. *British Journal of Educational Technology*, *52*(1), 112-124. https://doi.org/1.1111/bjet.12957

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020b). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, *36*(4), 502-513. https://doi.org/1.1111/jcal.12416

Vandermaas-Peeler, M. (2008). Parental guidance of numeracy development in early childhood. In O. Saracho & B. Spodak (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 277–290). Information Age Publishing.

Vargas-Munoz, J. E., Srivastava, S., Tuia, D., & Falcao, A. X. (2021). OpenStreetMap: Challenges and Opportunities in Machine Learning and Remote Sensing. *IEEE Geoscience and Remote Sensing Magazine*, *9*(1), 184–199. https://doi.org/10.1109/mgrs.2020.2994107

Vegas, E., Ziegler, L., & Zerbino, N. (2019). *How ed-tech can help leapfrog progress in education* (pp. 1–18). Center for Universal Education at The Brookings Institution. https://files.eric.ed.gov/fulltext/ED602936.pdf

Verdolini, E., Anadón, L. D., Baker, E., Bosetti, V., & Aleluia Reis, L. (2018). Future prospects for energy technologies: Insights from expert elicitations. *Review of Environmental Economics and Policy*, *12*(1), 133-153. https://doi.org/10.1093/reep/rex028

Walter-Laager, C., Brandenberg, K., Tinguely, L., Schwarz, J., Pfiffner, M. R., & Moschner, B. (2016). Media-assisted language learning for young children: Effects of a word-learning app on the vocabulary acquisition of two-year-olds. *British Journal of Educational Technology*, *48*(4), 1062–1072. https://doi.org/10.1111/bjet.12472

Walton, J. M. (2018). *Evaluating the impact of a tablet-based intervention on the mathematics attainment, receptive language and approaches to learning of preschool children* [Doctoral dissertation, University of Nottingham]. ResearchGate. https://www.researchgate.net/publication/329585578_Evaluating_the_impact_of_a_tablet-based_intervention_on_the_mathematics_attainment_receptive_language_and_approaches_to_learning_of_preschool_children

Wang, G., Zhang, Y., Zhao, J., Zhang, J., & Jiang, F. (2020). Mitigate the effects of home confinement on children during the COVID-19 outbreak. *The Lancet*, *395(*10228), 945-947. https://doi.org/10.1016/S0140-6736(20)30547-X

Webb, D., Barringer, K., Torrance, R., & Mitchell, J. (2020). *Girls' Education and EdTech: A Rapid Evidence Review*. EdTech Hub. https://doi.org/10.5281/zenodo.3958002

Webb, S., Holford, J., Hodge, S., Milana, M., & Waller, R. (2017). Lifelong learning for quality education: exploring the neglected aspect of sustainable development goal 4. *International Journal of Lifelong Education*, *36*(5), 509–511. https://doi.org/10.1080/02601370.2017.1398489

Wei, W., Wu, Y., Lv, B., Zhou, H., Han, X., Liu, Z., & Luo, L. (2016). The Relationship Between Parental Involvement and Elementary Students' Academic Achievement in China: One-Only Children vs. Children with Siblings. *Journal of Comparative Family Studies*, *47*(4), 483–500. https://doi.org/10.3138/jcfs.47.4.483

White, C. (2015). Rurality, urbanity, indigeneity and schooling in Fiji. *International Education*, *44*(2), 69.

Williams, F. E. (2022). Learning at Taonga Market. *LC Journal of Special Education*, *3*(15). https://digitalshowcase.lynchburg.edu/cgi/viewcontent.cgi?article=1043&context=lc-journal-of-special-education

Williams, M. J. (2018). External Validity and Policy Adaptation: From Impact Evaluation to Policy Design. *The World Bank Research Observer*, *35*(2), 158–191. https://doi.org/10.1093/wbro/lky010

Willoughby, D., Evans, M. A., & Nowak, S. (2015). Do ABC eBooks boost engagement and learning in preschoolers? An experimental study comparing eBooks with paper ABC and storybook controls. *Computers & Education*, *82*, 107–117. https://doi.org/10.1016/j.compedu.2014.11.008

Workman, J. (2017). Sibling Additions, Resource Dilution, and Cognitive Development During Early Childhood. *Journal of Marriage and Family*, *79*(2), 462–474. https://doi.org/10.1111/jomf.12350

World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise*. https://doi.org/10.1596/978-1-4648-1096-1

World Bank. (2019, January 22). *The Education Crisis: Being in School Is Not the Same as Learning.* https://www.worldbank.org/en/news/immersive-story/2019/01/22/pass-or-fail-how-can-the-world-do-its-homework

World Bank. (2021, April 18). *Keeping Bangladesh's Students Learning during the COVID-19 Pandemic – Results Brief.* https://www.worldbank.org/en/results/2021/04/18/keeping-bangladesh-s-students-learning-during-the-covid-19-pandemic

World Bank. (2022a). *Girls' Education*. https://www.worldbank.org/en/topic/girlseducation

World Bank. (2022b). *Low & middle income | Data*. https://data.worldbank.org/income-level/low-and-middle-income

World Bank. (2023, April 27). *World Bank Comparison: Curious Learning apps for early literacy learning were second only to one-on-one tutoring*. Curious Learning. https://www.curiouslearning.org/essays/2023/4/27/world-bank-comparison-curious-learning-apps-for-early-literacy-learning-were-second-only-to-one-on-one-tutoring

World Bank, Foreign, Commonwealth & Development Office (FCDO), & BE2. (2020). *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?* https://www.worldbank.org/en/topic/teachingandlearning/publication/cost-effective-approaches-to-improve-global-learning

World Bank, UNESCO, UNICEF, USAID, FCDO, & Bill & Melinda Gates Foundation. (2022). *The State of Global Learning Poverty: 2022 Update.* UNICEF. https://www.unicef.org/reports/state-global-learning-poverty-2022

World Health Organization. (2017). *Determinants of health*. https://www.who.int/news-room/questions-and-answers/item/determinants-of-health

World Health Organization. (2023, June 2). *Adolescent pregnancy*. https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy

World Health Organization. (2024, February 5). *Female Genital Mutilation*. https://www.who.int/news-room/fact-sheets/detail/female-genital-mutilation

Wright, L-A., & Plasterer, R. (2012). Beyond Basic Education: Exploring Opportunities for Higher Learning in Kenyan Refugee Camps. *Refuge: Canada's Journal on Refugees*, *27*(2), 42–56. https://doi.org/10.25071/1920-7336.34721

XPRIZE. (2017, September). *Empowering Young Minds Everywhere: Five Teams Advance To Final Round of $15M GLXP*. https://www.xprize.org/prizes/global-learning/articles/empowering-young-minds-everywhere-five-teams-advance

XPRIZE, (2019). *GLXP Executive Summary*. https://www.xprize.org/prizes/global-learning/articles/glexp-executive-summary

XPRIZE. (2020). *Home page* [GitHub repository]. https://github.com/XPRIZE

Yang, J. C., & Quadir, B. (2018). Individual differences in an English learning achievement system: gaming flow experience, gender differences and learning motivation. *Technology, Pedagogy and Education*, *27*(3), 351–366. https://doi.org/10.1080/1475939x.2018.1460618

Yardi, S., & Bruckman, A. (2012). Income, race, and class: Exploring socioeconomic differences in family technology use. *Proceedings of the 2012 ACM Annual Conference* *on Human Factors in Computing Systems - CHI '12*. https://doi.org/10.1145/2207676.2208716

Yu, R. (2020). Research on the Innovation of Personalized Education in Colleges and Universities under the Context of Big Data. *Journal of Contemporary Educational Research*, *4*(2), 77–81. https://doi.org/10.26689/jcer.v4i2.1017

Zacharia, S. (2020). *Education continuity during the Coronavirus crisis: Pakistan—TeleSchool and Taleem Ghar (Educational TV at Home).* World Bank. *https://documents1.worldbank.org/curated/en/421821600058352361/pdf/Pa kistan-TeleSchool-and-Taleem-Ghar-Educational-TV-at-Home.pdf.*

Zhang, F., Jiang, Y., Ming, H., Ren, Y., Wang, L., & Huang, S. (2020). Family socio-economic status and children's academic achievement: The different roles of parental academic involvement and subjective social mobility. *British Journal of Educational Psychology*, *90*(3), 561–579. https://doi.org/10.1111/bjep.12374

Zhao, D., He, Z., Tian, Y., & Liu, H. (2022). Differences in Cognitive and Non-Cognitive Results between Only-Child and Non-Only-Child Children: Analysis of Propensity Scores Based on Large-Scale Assessment. *Children*, *9*(6), 807. https://doi.org/10.3390/children9060807

Appendices

Appendix A: Search strategy used to identify out-of-school learning intervention studies for Chapter 1 - Introduction & Literature Review.

Search string

((Child* OR adolescen* OR teen* OR youth) AND (Out-of-school OR "out of school" OR "no* educat*") AND (Mobile tech* OR smartphone OR laptop OR tablet) AND (Intervention OR competition OR program* OR trial OR increas* OR decreas* OR address* OR change* OR impact OR prevent* OR support* OR improv*) AND (Learn* OR educat*))

Countries

LMICs were identified based on the World Bank classification, last updated in 2022 (World Bank, 2022b). This identifies 136 countries as eligible. Rather than incorporate them into the search string, studies from high-income countries were excluded during the title/abstract and full-text screening stages.

Search

Literature searches were conducted in the following:

1. Academic databases, including:
    - Scopus
    - ProQuest
    - ERIC
    - Web of Science
    - JSTOR
    - ScienceDirect
    - PsychINFO
    - DOAJ

- Project MUSE
- IEEE
- Child Development and Adolescent Studies
- IBSS
- Literature Online
- British Education Index

2. Research groups, including EdTech Hub, M-Education Alliance, The World Bank, RTI International.
3. NGO's, including UNESCO, WFP, UNICEF, Save the Children, Imagine Worldwide and the International Rescue Committee (IRC).
4. A grey literature search was conducted for difficult-to-locate studies, conference proceedings, PhD theses, policy documents, and research reports.

Appendix B: Contextual survey questions that the XPRIZE Foundation asked the children and their caregivers at the pre-test and/or post-test sessions for the GLXP competition

Select questions from the following were chosen to be engineered into features, based on what the literature base suggested may be important for learning (see Chapters 2 and 5 for further details).

Questions asked at baseline

Questions for the child:

How old are you?

Were you enrolled in school this past year?

If yes, in what grade?

If yes, how often did you go to school this past year?

If no, why didn't you go to school?

Did you eat any food before you arrived here today?

Apart from your school books, are there books, newspapers or other materials for you to read at home?

How often do you read out loud to someone at home?

How often does someone read to you at home?

Have you ever used a device like this before [tablet]?

Have you ever used a device like this before [smartphone]?

What language do you normally speak with your family at home?

Questions for the caregiver:

What is your relationship with the child?

Are you the head of the household?

If no, what is the relationship of head of household to the child?

How old is the participating child?

Do you know the child's birth month?

Do you know the child's birth year?

Has the child ever attended school?

If yes, in what type of school?

In what grade?

If no, what prevented the child from going to school?

What is the highest level of schooling you have reached?

Has anyone in your household completed a higher level of schooling than you?

If yes, specify level.

How many older siblings does the participating child have?

Among the older siblings, how many know how to read and write?

How many younger siblings does the participating child have?

Among the younger siblings, how many know how to read and write?

What language or languages would you say you read well?

Please describe the primary type of work you do to support your family.

What language do you normally speak in your household?

Does your family have electricity in your home?

Where do you normally get your water from at home?

How is food most often cooked at your home?

What type of toilet does your family use at your home?

Does your household have a:

Radio or cassette/tape recorder?

Mobile phone?

Lantern?

Table?

Television?

Bicycle?

Motorbike?

Cattle?

What is the main building material used for the walls of your home?

What is the main building material used for the roof of your home?

Have you ever used a device like this before [tablet]?

For further communication, can you give us a telephone contact?

Do you own the phone for the number you provided?

What type of phone do you have/use?

Questions asked at endline (all directed to the child):

How old are you?

Were you enrolled in school this past year?

If yes, in what grade?

Are you attending school now with other children?

If yes, what kind of school?

If yes, how often do you go to school/attend lessons?

If no, why didn't you go to school?

Does the village Mama help you use or learn from your tablet?

Does any adult in the home help you use or learn from your tablet?

Does any sibling in the home help you use or learn from your tablet?

Does anyone else help you use or learn from your tablet?

Apart from your school books, are there books, newspapers or other materials to use in your home?

How often do you read out loud to someone at home?

How often does someone read to you at home?

Since you got the tablet, how often do you use it?

Appendix C: List of variables reported in the primary XPRIZE dataset relating to the EGRA/EGMA tests.

The variables below, shown in Table A1, relate to the data collected at the baseline test. This has been shortened where there were many questions for the same theme (e.g. Syllable Sounds 1-100 - the repetition has been explained in brackets to indicate how many questions this covered).

Table A1

A table depicting the variables collected from the baseline EGRA/EGMA tests administered in 2017 during the GLXP competition

Variable Name	Variable Label
treat_alpha	Alphabetic Treatment Assignment
village_id	VILLAGE_ID
id	id
child_name_tablet	cq_child_name
pq_child_name_Base	pq_child_name
district	District
ward	Ward
village	Village
kitongoji	Kitongoji
new_child_Base	Child not at Baseline
enumerator_Base	Enumerator conducting assessment
start_time_Base	Test start time?
orf_Base	Oral Reading Fluency
csspm_Base	Correct Syllable Sounds Per Minute
cwpm_Base	Correct Words Per Minute
cnonwpm_Base	Correct Invented Words Per Minute
print1_Base	1. On this page, where would you begin to read? Please show me with your finger.
print2_Base	2. Now show me in which direction you would read next.

Variable Name	Variable Label
print3_Base	3. When you get to the end of the line, where would you read next?
syll_sound1_Base	Syllable Sound; Item 1 (repeated for items 1-100)
syll_sound_score_Base	Total correct Syllable Sound questions.
syll_sound_time_remain_Base	Time remaining when finished answering Syllable Sound questions?
syll_sound_score_pcnt_Base	Percentage of questions correct.
syll_sound_score_zero_Base	Student scored zero on section.
syll_sound_auto_stop_Base	Could the child not complete any questions?
syll_sound_attempted_Base	Number of questions attempted.
syll_sound_attempted_pcnt_Base	Percentage of attempted questions correct.
fam_word1_Base	Familiar Word; Item 1 (repeated for items 1-50)
fam_word_score_Base	Total correct Familiar Word questions.
fam_word_time_remain_Base	Time remaining when finished answering Familiar Word questions?
fam_word_score_pcnt_Base	Percentage of questions correct.
fam_word_score_zero_Base	Student scored zero on section.
fam_word_auto_stop_Base	Could the child not complete any questions?
fam_word_attempted_Base	Number of questions attempted.
fam_word_attempted_pcnt_Base	Percentage of attempted questions correct.

Variable Name	Variable Label
invent_word1_Base	Invented Word; Item 1 (repeated for items 1-50)
invent_word_score_Base	Total correct Invented Word questions.
invent_word_time_remain_Base	Time remaining when finished answering Invented Word questions?
invent_word_score_pcnt_Base	Percentage of questions correct.
invent_word_score_zero_Base	Student scored zero on section.
invent_word_auto_stop_Base	Could the child not complete any questions?
invent_word_attempted_Base	Number of questions attempted.
invent_word_attempted_pcnt_Base	Percentage of attempted questions correct.
oral_read1_Base	Oral Reading Word; Item 1 (repeated for items 1-42)
oral_read_score_Base	Total correct Oral Reading Word questions.
oral_read_time_remain_Base	Time remaining when finished answering Oral Reading Word questions?
oral_read_score_pcnt_Base	Percentage of questions correct.
oral_read_score_zero_Base	Student scored zero on section.
oral_read_auto_stop_Base	Could the child not complete any questions?
oral_read_attempted_Base	Number of questions attempted.
oral_read_attempted_pcnt_Base	Percentage of attempted questions correct.
read_comp_score_pcnt80_Base	Child understood 80% or more of the passage
read_comp_attempted_pcnt80_Base	Child understood 80% or more of what they read in the passage

Variable Name	Variable Label
read_comp1_Base	Reading Comprehension; Item 1
read_comp2_Base	Reading Comprehension; Item 2
read_comp3_Base	Reading Comprehension; Item 3
read_comp4_Base	Reading Comprehension; Item 4
read_comp5_Base	Reading Comprehension; Item 5
read_comp_score_Base	Total correct Reading Comprehension questions.
read_comp_score_pcnt_Base	Percentage of questions correct.
read_comp_score_zero_Base	Student scored zero on section.
read_comp_attempted_Base	Number of questions attempted.
read_comp_attempted_pcnt_Base	Percentage of attempted questions correct.
list_comp1_Base	Listening Comprehension; Item 1
list_comp2_Base	Listening Comprehension; Item 2
list_comp3_Base	Listening Comprehension; Item 3
list_comp4_Base	Listening Comprehension; Item 4
list_comp5_Base	Listening Comprehension; Item 5
list_comp_score_Base	Total correct Listening Comprehension questions.
list_comp_score_pcnt_Base	Percentage of questions correct.
list_comp_score_zero_Base	Student scored zero on section.
list_comp_attempted_Base	Number of questions attempted.
list_comp_attempted_pcnt_Base	Percentage of attempted questions correct.

Variable Name	Variable Label
dict_q1a_Base	[d] Please copy the letters on this page onto the paper that I have given you.
dict_q1b_Base	[m] Please copy the letters on this page onto the paper that I have given you.
dict_q1c_Base	[e] Please copy the letters on this page onto the paper that I have given you.
dict_q2_Base	Please copy the word on this page onto the paper that I have given you. [kapu]
dict_q3_Base	Please write the sentence: Nyumbani kwenu ni wapi? (Where do you live?)
dict_q4_Base	Please use your own words to write a short response to the question that you have.
write_photo_captured_Base	Writing Subtask Photo Captured?-Baseline
write_photo_url_Base	Writing Subtask URL- Baseline
caddpm_Base	Correct Addition Problems Per Minute
csubpm_Base	Correct Subtraction Problems Per Minute
cnumidpm_Base	Correct Number Identification Problems Per Minute
curr1_Base	1. Here are two piles of coins. Please show me which pile has more coins in it.
curr2_Base	2. Here are some Tanzanian banknotes. Please tell me how many bank notes there are

Variable Name	Variable Label
curr3_Base	3. Please tell me how many shillings those banknotes represent altogether.
curr4_Base	4. Please tell me how many shillings those banknotes represent altogether.
curr5_Base	5. Here is a picture of something that I want to buy. This is the price [point t
curr_score_Base	Currency Score
curr_score_pcnt_Base	Currency Percent Score
curr_score_zero_Base	Did Child Score Zero on Currency Subtask?
curr_attempted_Base	Number of Items Attempted
curr_attempted_pcnt_Base	Percent Correct of Attempted
num_id1_Base	Number Identification; Item 1 (repeated for items 1-20)
num_id_score_Base	Total correct Number Identification questions.
num_id_time_remain_Base	Time remaining when finished answering Number Identification questions?
num_id_score_pcnt_Base	Percentage of questions correct.
num_id_score_zero_Base	Student scored zero on section.
num_id_auto_stop_Base	Could the child not complete any questions?
num_id_attempted_Base	Number of questions attempted.
num_id_attempted_pcnt_Base	Percentage of attempted questions correct.

Variable Name	Variable Label
quant_comp1_Base	[7,5] Look at these numbers. Tell me which number is bigger.
quant_comp2_Base	[16, 23] Look at these numbers. Tell me which number is bigger.
quant_comp3_Base	[39, 23] Look at these numbers. Tell me which number is bigger.
quant_comp4_Base	[1/3, 1/4] Look at these numbers. Tell me which number is bigger.
quant_comp5_Base	[65, 67] Look at these numbers. Tell me which number is bigger.
quant_comp6_Base	[77, 67] Look at these numbers. Tell me which number is bigger.

The endline questions are not included as the exact same tests were administered and variables collected. The following variables were provided in the data that measure differences between the baseline and endline tests, as shown in Table A2.

Table A2

A table depicting the variables engineered from the baseline-endline EGRA/EGMA tests administered in 2017 and 2019 for the GLXP competition

Variable Name	Variable Label
syll_sound_score_pcnt_gain	syll_sound_score_pcnt Baseline to Endline Gain
fam_word_score_pcnt_gain	fam_word_score_pcnt Baseline to Endline Gain
invent_word_score_pcnt_gain	invent_word_score_pcnt Baseline to Endline Gain
oral_read_score_pcnt_gain	oral_read_score_pcnt Baseline to Endline Gain
read_comp_score_pcnt_gain	read_comp_score_pcnt Baseline to Endline Gain
list_comp_score_pcnt_gain	list_comp_score_pcnt Baseline to Endline Gain
num_id_score_pcnt_gain	num_id_score_pcnt Baseline to Endline Gain
quant_comp_score_pcnt_gain	quant_comp_score_pcnt Baseline to Endline Gain
miss_num_score_pcnt_gain	miss_num_score_pcnt Baseline to Endline Gain
word_prob_score_pcnt_gain	word_prob_score_pcnt Baseline to Endline Gain
addlvl2_score_pcnt_gain	addlvl2_score_pcnt Baseline to Endline Gain
add_score_pcnt_gain	add_score_pcnt Baseline to Endline Gain
sublvl2_score_pcnt_gain	sublvl2_score_pcnt Baseline to Endline Gain
sub_score_pcnt_gain	sub_score_pcnt Baseline to Endline Gain
syll_sound_score_zero_gain	syll_sound_score_zero Baseline to Endline Gain
fam_word_score_zero_gain	fam_word_score_zero Baseline to Endline Gain
invent_word_score_zero_gain	invent_word_score_zero Baseline to Endline Gain
oral_read_score_zero_gain	oral_read_score_zero Baseline to Endline Gain
read_comp_score_zero_gain	read_comp_score_zero Baseline to Endline Gain
list_comp_score_zero_gain	list_comp_score_zero Baseline to Endline Gain

Variable Name	Variable Label
num_id_score_zero_gain	num_id_score_zero Baseline to Endline Gain
quant_comp_score_zero_gain	quant_comp_score_zero Baseline to Endline Gain
miss_num_score_zero_gain	miss_num_score_zero Baseline to Endline Gain
word_prob_score_zero_gain	word_prob_score_zero Baseline to Endline Gain
addlvl2_score_zero_gain	addlvl2_score_zero Baseline to Endline Gain
add_score_zero_gain	add_score_zero Baseline to Endline Gain
sublvl2_score_zero_gain	sublvl2_score_zero Baseline to Endline Gain
sub_score_zero_gain	sub_score_zero Baseline to Endline Gain
syll_sound_score_gain	syll_sound_score Baseline to Endline Gain
fam_word_score_gain	fam_word_score Baseline to Endline Gain
invent_word_score_gain	invent_word_score Baseline to Endline Gain
oral_read_score_gain	oral_read_score Baseline to Endline Gain
read_comp_score_gain	read_comp_score Baseline to Endline Gain
list_comp_score_gain	list_comp_score Baseline to Endline Gain
num_id_score_gain	num_id_score Baseline to Endline Gain
quant_comp_score_gain	quant_comp_score Baseline to Endline Gain
miss_num_score_gain	miss_num_score Baseline to Endline Gain
word_prob_score_gain	word_prob_score Baseline to Endline Gain
addlvl2_score_gain	addlvl2_score Baseline to Endline Gain
add_score_gain	add_score Baseline to Endline Gain
sublvl2_score_gain	sublvl2_score Baseline to Endline Gain
sub_score_gain	sub_score Baseline to Endline Gain
orf_gain	orf Baseline to Endline Gain

Variable Name	Variable Label
csspm_gain	csspm Baseline to Endline Gain
cwpm_gain	cwpm Baseline to Endline Gain
cnonwpm_gain	cnonwpm Baseline to Endline Gain
caddpm_gain	caddpm Baseline to Endline Gain
csubpm_gain	csubpm Baseline to Endline Gain
cnumidpm_gain	cnumidpm Baseline to Endline Gain