

REPORT

Estimating the Impact of Educational Television on Literacy, Gender, and SEL



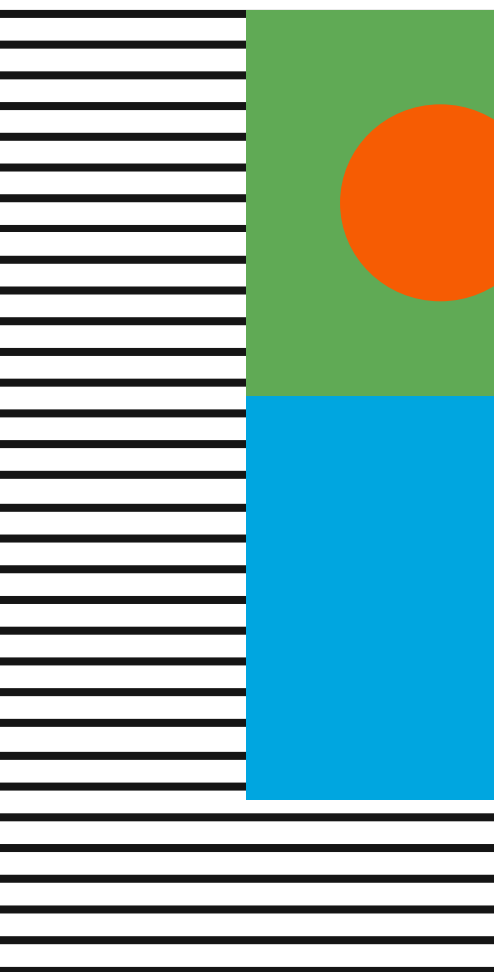


How to cite

Katumbi, Lara; Salome, Njambi; Anushka, Ghosh; Shilvaan, Patel; and Nicolas Bottan. *Estimating the Impact of Educational Television on Literacy, Gender, and SEL*. Busara, Ubongo EdTech Endline Report. Nairobi: Busara, 2024. DOI: doi.org/10.62372/GKQN1155

Contents

4	Introduction
	Theory of Change
	Intervention Overview
	Episode Synopsis
13	Study Design and Methodology
	Research Questions
	Outcome Measures
	Methodology
	Data collection
	Analysis Approach
19	Results
	Descriptive Statistics
	Compliance
	Treatment Effects on Main Outcomes
	Local Average Treatment Effects
	Heterogeneity
	Mechanisms
	Cost Effectiveness Analysis
40	Learnings and Recommendations
45	References



1

Introduction



Introduction

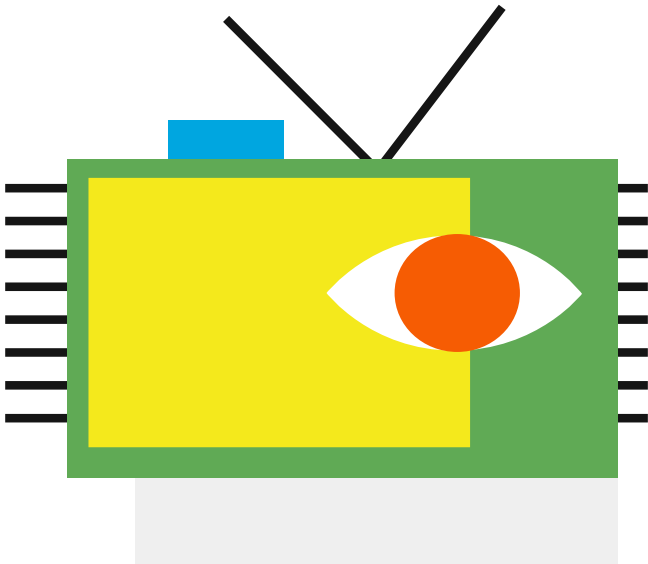
Although access to primary education has increased dramatically in the last two decades, recent attention has focused on “the learning crisis”. In 2019, 57% of children in low- and middle-income countries (LMICs) and 86% in Sub-Saharan Africa did not achieve the minimum reading proficiency by age 10 (WorldBank et al., 2022). The COVID-19 pandemic has exacerbated this phenomenon as an estimated 1.2 billion children in 186 countries were affected by school closures at the height of lockdown measures (UNESCO, 2022; Tadesse and Muluye, 2020). Addressing these shortfalls in the short and medium term is difficult in light of the dearth of key resources including lack of internet, quality teachers, and educational technology (EdTech) relating to educational television (TV) (UNESCO, 2022). While interventions such as computer-aided learning have been shown to be highly effective at improving learning they come at a high cost especially in contexts that require investment in additional infrastructure (e.g., Muralidharan et al., 2019; Araya et al., 2019). EdTech including broadcasting of educational TV on public channels can play an important role in bridging the educational gap even in contexts with restricted access to the Internet and formal schooling infrastructure. However, there are limited studies investigating the impact of EdTech including watching educational TV shows in a naturalistic setting on educational outcomes in the context of LMICs in Sub-Saharan Africa.

This paper presents evidence from a large randomized controlled trial (RCT) in Kenya, investigating the effects of watching a new children’s educational TV show at home. One novel aspect of the show is that, besides being instructional, it focuses on changing children’s mindsets about reading, gender attitudes, and socio-emotional learning. We recruited 4,300 children from 346 public schools. Students in randomly selected

treatment schools received encouragement to watch the publicly broadcasted TV show. We sent parents biweekly SMS reminders about the time and channel the TV show aired. Students in control schools did not receive any encouragement or reminders about the show. The TV show is a cartoon program named Nuzo & Namia that targets children aged 6 to 9. It was created in 2023 by Ubongo and its main educational objective is to improve literacy. Additionally, it aims to change gender norms and encourage different forms of socio-emotional learning (i.e., confidence and curiosity). Identifying the causal effect of our encouragement treatment relies on random assignment at the school level. Our primary estimates captured intention-to-treat (ITT) effects. We used a household survey to collect data. We collected endline data a year after the baseline and after about 9 months of exposure to the TV show. The endline survey also collected self-reported measures of watching the show to estimate the local average treatment effects (LATE) for treatment compliers.

We study three categories of main outcomes: literacy including reading and comprehension, gender attitudes, and socio-emotional learning (SEL). We employed the Early Grade Reading Assessment Tool (United States Agency for International Development [USAID], 2016) to measure literacy and construct a measure for reading fluency and comprehension. We study three dimensions of gender attitudes: gender stereotype knowledge and flexibility, gender roles, and in- and out-group attitudes and activities.. As a measure of socio-emotional learning, we construct a score combining confidence and curiosity measures. The ITT estimates from the encouragement design reveal that the Nuzo and Namia TV show had no effect on literacy. However, we find a positive and statistically significant impact of the show on gender attitudes and on SEL.

Following La Ferrara (2016), we considered three main mechanisms through which educational TV can drive changes in outcomes. First, the show can provide



viewers with new information and skills that can lead to better literacy either directly (e.g., the show teaches key strategies in early reading comprehension) or indirectly (e.g., learning strategies make reading easier or more enjoyable, thus increasing practice). Second, the show can mold viewers' preferences, such as having a positive effect on reading. This could encourage children to read more, thus improving literacy. There are additional behavioral mechanisms associated with changing preferences, such as demanding more books and reading with parents, among others. Third, outcomes can be affected through changes in time use where encouraging children to watch the show can crowd in or crowd out reading time or other behavior (e.g., watching more educational or non-educational TV [\[3.6\]](#)

We also considered indirect mechanisms from our encouragement treatment. For example, friends can watch the show and be exposed to the aforementioned mechanisms. In turn, social influence could increase both watching the show and reading behavior. Similarly, teachers could incorporate aspects of the show in their teaching (e.g., examples or better explanations), improving teaching effectiveness. Additionally, because our encouragement treatment is addressed to parents or caregivers, this could change their perception of the importance of reading, increasing their involvement in their child's education and educational activities. [\[3.6.2\]](#)

This study also explored the importance of different characteristics of the show. These can affect outcomes in two ways. The first is through engagement watching the show (i.e., treatment intensity). The second relates to mediators driving changes in children's beliefs and attitudes. Following social learning theory (**Bandura and Walters, 1977**), we explore the role played by features such as how relatable characters are to children that are important in adopting norms and behaviors. [\[3.6.3\]](#)

We also tested the role of mediating factors in driving the mechanisms. For instance, a qualitative study conducted as part of this evaluation found that the presence of a caregiver while watching the show played an important role in maintaining the child's engagement with the show. Additionally, if the show fosters reading, the presence of reading materials in the household is likely to positively or negatively moderate the effect of the show on the outcomes of interest. [\[3.6.4\]](#)

Learning outcomes in Kenya have generally been poor, with approximately 3 out of 10 children in grade 3 only being able to perform grade 2 level literacy tasks (Uwezo, 2016). Such underperformance, often the result of insufficient cognitive stimulation and nutrition (Baker-Henningham & Boo, 2010; Walker, 2011), can exacerbate life-long learning inequities of achievement and foster intergenerational transmission of poverty, thereby reducing the efficiency of public spending on primary and secondary education (Engle, et al., 2007; Grantham-McGregor, et al., 2007; Naudeau et al., 2011).

Overall, TV access has been growing steadily in LMICs (Center, 2018). About 45% of households in Kenya own a TV (Uwezo, 2020). Leveraging on this, Educational television offers a promising complement or alternative to conventional modes of instruction because it does not require teacher guidance, is low-cost, scalable, and thus has potential for reaching millions of children.

In response to this, the Government of Kenya has implemented dramatic shifts in basic education by implementing the Competency Based Curriculum (CBC). The new curriculum is rooted in the promotion of individual wellbeing and acquisition of skills and capabilities (Akala, 2021). The Government of Kenya has also increased government expenditure in the education sector (KNBS, 2021) and launched several scaled national policies aimed at improving literacy and numeracy, such as the Tusome National Literacy Program and the Primary Mathematics and Reading (PRIMR) Initiative.

Notwithstanding, children in Kenya faced significant barriers in accessing quality education due to limited access to EdTech and in particular educational television programs that are context specific. There was a surge in new forms of EdTech including language apps, virtual tutoring, video conferencing tools, and online learning software during the COVID-19 pandemic (World Economic Forum, 2020) and EdTech became the primary source of education for children (Market Insider, 2020). Overall, TV access has been growing steadily in LMICs (Center, 2018). About 45% of households in Kenya own a TV (Uwezo, 2020). Leveraging on this, Educational television offers a promising complement or alternative to conventional modes of instruction because it does not require teacher guidance, is low-cost, scalable, and thus has potential for reaching millions of children. Therefore, a natural scale-up with the potential of reaching millions would involve a


shift from watching educational shows in classrooms to households.

A growing body of literature suggests that educational television can improve children's learning outcomes in LMICs. A meta-analysis of laboratory and quasi experiments including 16 Sesame Street evaluations in LMICs suggests that the show had a positive impact on children's learning outcomes (Mares and Pan, 2013). Beyond Sesame Street, Akili and Me and Ubongo Kids, two programs produced by Ubongo show similar impacts on learning outcomes in LMICs in Sub-Saharan Africa. Quasi-experimental studies investigating the impact of Akili and Me—which targets children aged 3 to 6—show that the program had a positive impact on literacy, numeracy and socio-emotional skills including emotion identification among children in Tanzania and Rwanda (Borzekowski et al., 2018, 2019, 2020). In Kenya, an exploratory study finds that Ubongo Kids has a positive effect on children's mathematics capability (Watson et al., 2021). Furthermore, an experimental study targeting children aged 10 to 14 in Tanzania shows that Ubongo Kids significantly improved social emotional mindsets and skills i.e., curiosity, teamwork, growth mindset growth among others, and gender equity outcomes (Cherewick et al., 2021). In Haiti, an experimental study finds that educational TV including counter-stereotypical gender messaging diminishes beliefs surrounding gender stereotypes thus improving children's gender perceptions (Borzekowski et al., 2024).

Whilst existing studies provide empirical evidence that educational television can significantly improve learning and socio-emotional outcomes and gender attitudes in LMICs in Sub-Saharan Africa, they are not without limitations. Studies estimating causal effects rely on relatively small sample sizes (~300-600 children), narrow geographic regions within countries, and short intervention timeframes (~2-12 weeks). As such, a large evaluation is much needed to demonstrate whether these effects hold in naturalistic at-scale contexts.

The current study makes several contributions. First, this study uses a large RCT covering expansive geographical regions in Kenya to investigate the causal impact of educational TV on literacy, gender attitudes, and SEL. Past studies in Kenya have relied on descriptive designs only. Second, this study develops and validates a novel instrument to capture gender attitudes among children. There are limited validated measures for capturing gender attitudes among children. Third, this study contributes to existing literature on how TV shows and representation in the media can change social norms and behaviours among children. In particular, this study contributes to this literature by providing novel experimental evidence on TV shows' effectiveness in shaping children's social

norms – particularly gender attitudes. Fourth, the current study investigates how the impact of educational TV varies by children's gender. Fifth, this study investigates direct and indirect mechanisms through which educational TV drives changes in the outcomes of interest. Lastly, we also investigate mediating factors that moderate the effect of the mechanisms.



A meta-analysis of laboratory and quasi experiments including 16 Sesame Street evaluations in LMICs suggests that the show had a positive impact on children's learning outcomes (Mares and Pan, 2013). Beyond Sesame Street, Akili and Me and Ubongo Kids, two programs produced by Ubongo show similar impacts on learning outcomes in LMICs in Sub-Saharan Africa. Quasi-experimental studies investigating the impact of Akili and Me—which targets children aged 3 to 6

1.1

Theory of Change

The theory of change behind the encouragement treatment for learning in our study examines various mechanisms and barriers that could influence the effectiveness of the educational TV show “Nuzo & Namia” on children’s literacy, gender attitudes, and socio-emotional learning. The initial stage of our study posits that receiving encouragement increases the likelihood and intensity of watching the TV show. However, barriers such as limited access to a television and the viewing environment, including the presence of caregivers, could affect this stage. Our hypothesis is that watching the TV-show can improve literacy through both direct and indirect effects, such as reading more frequently (**Merga and Roni, 2017**), as well as through changes in preferences, information acquisition, and time use (**La Ferrara, 2016**).

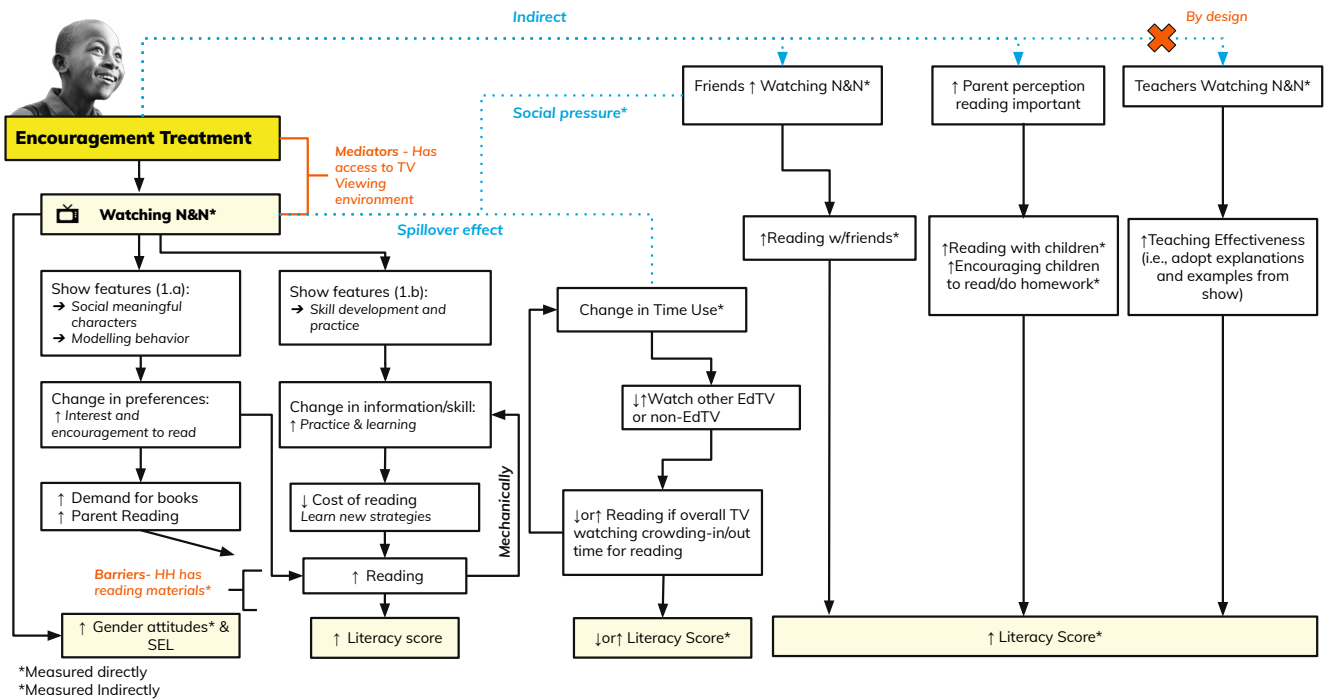
One primary mechanism is changing children’s preferences for reading. The show aims to engage children and increase their interest in reading by presenting socially meaningful characters and modelling behaviors that promote reading. This change in mindset could lead to higher motivation to read, demand for books, and parental involvement in reading activities. However, a significant barrier could be the lack of reading materials at home, limiting the effectiveness of increased reading motivation. Besides literacy, the show also aims to influence gender attitudes and socio-emotional learning, which can have reciprocal effects on literacy (**Alan et al., 2018, 2019; Yu et al., 2023; Deole and Zeydanli, 2021**).

The second mechanism involves directly transmitting information and skills related to reading and comprehension through the show. By teaching key reading strategies, the show can improve literacy both directly by applying these skills and indirectly by encouraging more reading practice. This could include interacting with

characters on the show or applying newly learned skills to other reading activities. Reduced reading difficulty due to these new skills could also lead to increased off-screen reading.

The third mechanism pertains to changes in children’s time use from watching the show. Watching “Nuzo & Namia” could either crowd in or crowd out other activities, including other TV watching or reading time. The net effect on literacy will depend on whether the show replaces non-educational content or displaces reading time. Parental involvement could also increase due to the encouragement treatment, with parents investing more in educational inputs and activities. By design, we are minimizing a potential channel that operates in other similar types of educational TV interventions conducted at school. Teachers typically play a central role in these interventions, facilitating viewing and discussions. Teachers could also benefit from watching the show by adopting examples and explanations from it in their teaching. This could increase their effectiveness in the classroom as research has found that providing teachers with better teaching plans and materials improves learning (**International Institute for Educational Planning (IIEP) - UNESCO, 2024**). However, our study takes several measures to shut down this channel. First, the recruitment and encouragement design operates exclusively outside of the school. Teachers were not involved at any step of the outreach or informational sessions. Second, also by design, the show is not widely advertised or communicated.

THEORY OF CHANGE



1.2

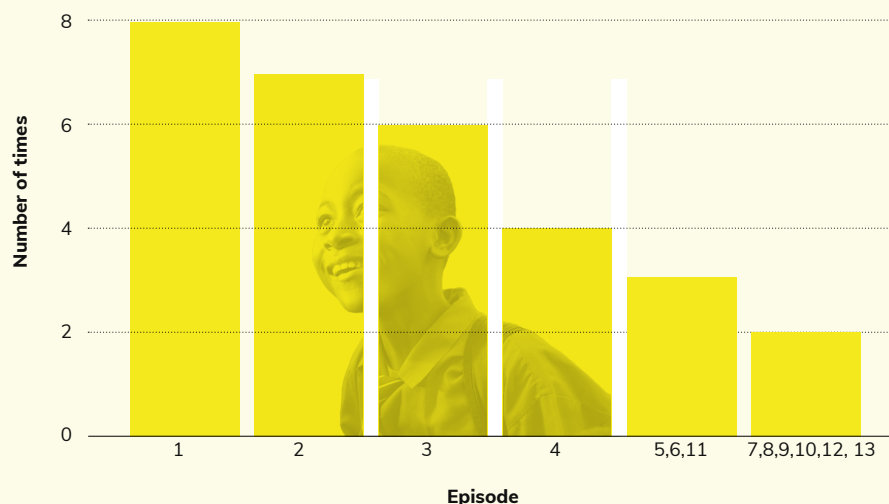
**NUZO & NAMIA
Intervention Overview**



The Nuzo and Namia show, developed by Ubongo, targeted 6-9-year-old children with the aim of improving their literacy, socio-emotional learning, and gender attitudes. The plot revolves around 2 protagonists, Nuzo and Namia, as they meet with different African characters at different moments throughout history. By interacting with the protagonists, who appear just in time to help them solve a problem that they are facing, viewers see and hear the protagonists model key strategies in early reading comprehension, such as predicting, questioning, clarifying, and analyzing information. For example, a character may predict a book’s content based on the story’s cover and title. The audience can then later see them refine their predictions halfway through the story or when they start reading. They may also compare texts they have read in different parts of the book and then self-question their comprehension.

Furthermore, seeing protagonists who are just like the audience (i.e., relatable role models) reading and using stories/books to practise their reading aims to give children agency, confidence, and motivation to improve

Overview of the number of times each episode was aired



and practise their own reading skills. In interactions with the historical figures in the show, along with their family and friends, viewers witness the protagonists identify and model positive socio-emotional learning outcomes and gender attitudes. The protagonists are taken on adventures where they can try things that earlier may have seemed unsuitable or impossible beforehand. Through the open-mindedness and perseverance of the protagonists, viewers learn that by trying different things and sticking to them for a while, they can find activities that truly suit them and their skills, even if societal norms or early difficulties don't encourage it. The series aims to both "show" these character traits (like confidence and curiosity) and explicitly state how these helped the protagonists succeed (e.g., a protagonist saying, "I am really glad I didn't give up when XY happened!").

The programme consists of 13 22-minute episodes. Each episode was broadcast twice a week for two weeks (i.e., four broadcasts per episode over a two-week period). The figure below gives an overview of the number of times each episode was aired.

We believe that this educational television program is valuable to our study for two reasons. First, the content focuses on early-grade literacy (comprehension), socio-emotional learning, and gender attitudes. Second, the program focuses on building foundational skills that align with and complement the national curriculum in Kenya. The Ubongo team adopted a rigorous approach to do so by identifying key elements of the national curriculum (e.g., summarizing), mapping specific learning outcomes for these elements, and developing content to support these learning outcomes. This allows us to understand how educational television can support curriculum-based learning. An overview of Ubongo's curriculum that we used in SMS engagement is illustrated in table 12.

1.3

Episode synopsis

We conducted a short observational ethnographic study at the midpoint of the intervention to gain insights on how children engage with the TV show. This section provides a snapshot of what a 30-minute Nuzo and Namia episode looks like, specifically focusing on episode 2.

Episode 2 of the show was centred around the key learning objective of developing skimming skills. The episode employs a playful and interactive approach to enhance the children’s comprehension strategy of skimming through various textual sources, such as newspaper articles, tour magazines, and brochures, to extract valuable information.

The show creatively incorporated hopscotch as a playful and engaging activity to reinforce skimming skills. As the characters embarked on their journey

to an unfamiliar area, they used newspaper articles, tour magazines, and brochures containing images and textual clues. Each frame on the hopscotch board represented a piece of information, and the characters had to skim through the provided materials to fill in the frames with relevant images.

Throughout the episode, the characters demonstrated diverse ways of interpreting the clues they encountered. This variation in interpretation highlights the importance of understanding different perspectives and encourages the development of social and emotional skills, such as confidence and curiosity. Despite their distinct interpretations, the characters ultimately reached the same destination, promoting teamwork and the value of diverse perspectives. An overview of what happens in the episode is illustrated in the table below where we took snapshots of the activities in the episode. It is worth noting that there were instances where the episode was not airing at all, due to the episode concluding earlier than the anticipated 30-minute duration.

Table 1: Overview of episode snapshots

Value Count	9	9	12	9
	One/more characters are reading out loud	A child is encouraged to practise a skill	One/more characters are giving instructions	Episode not playing at all
	13	6	23	
	One/more characters are demonstrating a skill	The child is encouraged to copy something	One/more characters are discussing something	

2

Study Design and Methodology



2.1

Research Questions

In this section, we address each evidence gap by exploring the following research questions in the study.

Primary

Estimating impact:

What is the impact of educational television on children's literacy, socio-emotional learning, and gender attitudes?

How does the impact vary by children's gender?

Costing impact:

What is the cost-effectiveness of educational television when children watch on television?

Secondary

What are the key mechanisms through which educational television impacts children's literacy, socio-emotional learning, and gender attitudes?

How does group viewing with a peer or a mentor affect the mechanisms through which literacy, socio-emotional learning, and gender attitudes are impacted?

How does supplementary educational television contribute to children's curriculum-based learning and learning experience in formal schooling? How does this vary by gender?



2.2

Outcome Measures

We have three main categories of primary outcomes: literacy, gender attitudes, and socio-emotional learning. Where applicable we denote secondary outcomes within those categories below.

1. Literacy Outcomes: Our literacy module was designed following the Early Grade Reading Assessment Tool (USAID, 2016). The instrument includes sub-sections for letter identification, non-word reading, oral fluency, and oral and listening comprehension. We construct two primary outcomes to measure literacy:

Reading Fluency Score: Aggregates scores for letter identification, non-word reading, and oral fluency.

Secondary Outcomes: We will examine the treatment effect on each sub-category that comprises the Reading Fluency Score (i.e., letter identification, non-word reading, etc.).

Comprehension Score: Aggregates scores for oral and listening comprehension. Secondary Outcomes: We will examine the treatment effect on each sub-category that comprises the comprehension score (i.e. Oral and Listening Comprehension).

2. Gender Attitudes: This study contributes by developing and validating measures to capture gender attitudes among children. We construct three main measures to measure different dimensions of gender attitudes.

Gender stereotype knowledge and flexibility score. Each child is asked whether they would typically see a boy or a girl performing various actions and exhibiting various characteristics. Following a mapping of which stereotypes are considered “male” or “female” in the literature, responses are coded with “1” if they follow the stereotype and “0” otherwise. These items are then aggregated into one score, denoting the number of

gender stereotypes the child subscribes to. A higher score means the child has strong gender stereotype knowledge and low stereotype flexibility.

Gender roles: An aggregate score measuring belief in traditional gender roles. The score aggregates individual questions about different behaviors and traits typically associated with either “male” or “female.” Each child is asked how likely they are to perform each behavior and how often they actually perform it. Based on each child’s gender and responses, a final score is constructed denoting the strength of their gender role beliefs. A higher score represents stronger beliefs in traditional gender roles.

Secondary Outcome: We will explore the score separately for female and male gender roles

In- and Out-group attitudes: An aggregate score measuring attitudes toward others with the same and different genders. The score aggregates individual questions about activities with both the in and out-group. Each child is asked how willing they are to perform each behavior and how often they actually perform it. Based on each child’s gender and responses, a final score is constructed to denote the strength of in-group and out-group attitudes. A higher score represents a strong preference for in-groups, while a lower score represents flexibility between in and out-groups.

Secondary Outcome: We will explore the score for in-group attitudes and activity for females and males separately.

3. Socio-emotional Learning: Confidence and Curiosity: We use measures of confidence and curiosity as a proxy for self-efficacy using the RTI confidence and curiosity scale, validated in both our East African context and our age group (Jukes et al., 2021). The score is the simple sum of all items, where relevant items are reverse-coded so that a higher final score reflects higher confidence and curiosity. Our main outcome here, defined as the “SEL Score” pools the scores for Confidence and Curiosity.

4. Secondary Outcomes: We will study the effects on confidence and curiosity scores separately as a secondary outcome.

2.3

Methodology

Sampling

Sample Selection

We conducted a school-level cluster randomized controlled trial. Our sampling strategy consisted of 3 stages: (1) identifying eligible counties and sub-counties, obtaining a random draw; (2) identifying eligible schools within randomly selected sub-counties and obtaining a random draw; (3) obtaining a final sample of children within eligible schools.

Selecting counties and sub-counties

We used data on public primary school enrollment from the Kenya Basic Education Statistical Booklet 2019 (Kenya Ministry of Education, 2019) and data from the Kenyan Census to obtain sub-county school-level estimates of eligible children. We defined an eligible child as one enrolled in grades 1-3 and with a TV at home. We then kept sub-counties where we estimated at least 40 eligible children enrolled in a public primary school and further dropped sub-counties that were the only eligible units within their county to reduce enumerator travel costs. We also excluded Nairobi because it is the capital city, as there are many schools in close geographic proximity, making spillovers more likely. We made the final selection of sub-counties by first specifying all possible combinations of 6 counties from the pool of eligible counties identified above, such that each county is in a different region. Each possible combination yields a list of at least 3,000 government primary schools. Finally, we randomly selected one of the combinations of 6 counties, yielding a list of over 3,000 government primary schools within their eligible sub-counties.

Selecting schools within sub-counties

We followed the same procedure used to select sub-counties to select schools. We first calculated the expected number of eligible boys and girls (i.e., in grades 1-3 and, potentially, with a TV at home). We then defined an eligible school as one in which we had estimated at least 12

eligible girls and 12 eligible boys. Kenya is divided into 47 counties and 314 sub-counties. Although we aimed at surveying 15 eligible students in each school, we targeted schools with more than 40 eligible children because (1) these are based on rough back-of-the-envelope estimates; (2) participation was voluntary, and parents might opt their child out of the study. There are 8 regions in Kenya. We targeted combinations of counties yielding at least 3,000 schools because we aimed to include 500 schools for the study that were at least 4 km away from each other. However, schools tend to be quite close together, so we used a conservative factor of 6 when defining the size. We used a threshold of 12 to ensure there was a large enough sample of children with access to TV. From this group of eligible schools, we randomly sampled schools such that each additional randomly sampled school was at least 4 kilometres away from any other schools already included in our sample.

Sampling children within these schools

Once we had our randomly selected schools, we obtained samples of children within these schools. Field officers were sent to the locations of our sampled schools. With the help of the village chief, the village elders, and the teachers, we obtained household contact information. We then conducted a phone screening and recruitment exercise to ensure that we had many eligible households that met our study criteria and were included in the study. These were interviewed during baseline. The criteria included children's age and TV access. Due to a fairly low number of households without access to a functioning TV, we included all eligible households in our sample. We also included children in the same household in our sample.

Randomization and Treatment Assignment

At baseline, we had originally designed the study to constitute two treatment arms but then decided to stick to one treatment arm due to challenges encountered during sampling. The treatment arm is as follows: TV schools: Group of schools that were provided with an encouragement to watch only the TV broadcast of the show (Treatment A) and the control group conducted “business as usual” and received no encouragement or information to watch the show as illustrated in **table 2**.

Creating treatment schools from control schools

To generate random variation in whether children watched the programme, we used an encouragement design. We randomly assigned 19 schools to be Control Schools or The Treatment (A) schools and encouraged eligible children within the latter to watch 2 episodes of the assigned programme per week for approximately 7 months. The encouragement was done through the children’s parents.

Parent-level encouragement approach

We held a meeting in each school at a local community center and invited parents of the selected students. We shared an overview of the show and how the show is positioned to support some of the children’s learning issues. We leveraged on some of the parents who knew about the show to share their opinions on the show content. To increase buy-in amongst the parents we distributed an

Ubongo factsheet, shared testimonials from other parents, and showed an example of a scene from Ubongo.

We expected our encouragement to randomly vary the ‘dosage’ of the programme between the schools, such that children in the Treatment Schools watch the programme significantly more than children in the Control Schools. We chose to use an encouragement design as we were unable to restrict access to the programme since it was distributed nationally on free-to-air television.

The encouragement of parents was conducted in several steps:

1

Committing to encourage the children to watch the show

Parents were asked to commit to encouraging their kids to watch the show on the days the show was aired by signing a commitment form

2

SMS reminder and quiz system

The parents were sent a reminder SMS, 30 minutes before the Ubongo show aired. After the show, they received an SMS quiz about the episode which had just aired.

3

Take-home material (pamphlet and poster)

Lastly, the parents were provided with an information pamphlet that contained:
Information on the time the show aired.
Instructions on how to access the channel from their television sets
A tracker for the parents to tick after their children watch an episode of the programme.

2.4

Data collection

Data was collected in two phases:

Baseline: before the launch of the intervention, approximately from April 2023- May 2023

Qualitative study: a month after the intervention launch in July 2024

Endline: Data collection took place immediately after the entire season of the educational television programme was aired. It happened over two weeks between 13 April 2024 and 26 April 2024.

Study Timelines

Activity	Start	Finish
Baseline Data collection	1-May-23	29-May-23
Launch of the encouragement	29-May-23	03-May-24
The Nuzo and Namia show airs	1-June-23	16-March- 24
Qualitative Study	10-August- 23	17-August-23
Conducted endline	13-April-24	26-April-24

2.5

Analysis Approach

Data was collected in two phases:

Our main analysis estimates our intervention’s intention-to-treat (ITT) effects on the treatment group. Because we employ a randomized school-level clustered-treatment approach, we can identify the causal effect by estimating the following model:

$$y_{is} = \alpha_0 + \alpha_1 \text{Treats} + \beta_1 y_{is,t=0} + \gamma_1 G_s + \gamma_2 X_s + \epsilon_{is}$$

Where y_{is} is an outcome for the child i in school s , Treats is an indicator for school s being treated.

We estimate 2 types of average treatment effects: (i) intent-to-treat (ITT) and (ii) complier average causal effect (CACE) (Gerber and Green, 2012). ITT measures the average effect of assignment to a treatment, regardless of

whether treatment is received. It then ignores compliance issues, such as whether the program or Control children actually watch the program. On the other hand, CACE measures the average effect of receipt of treatment (i.e., watching the program) among the subset of children who comply with their assigned program.

We include the baseline value of the outcome variable $y_{is, t=0}$, when available, to improve statistical precision, set to the mean if missing, and an indicator of missingness (Mis). G is an indicator for the school quadruplet during randomization, and X is a vector of school-level controls that includes the characteristics used in our stratified randomization.

Our primary parameter of interest is α_1 , which captures the direct effect of the treatment on children, as well as any potential spillovers on treated children within treatment schools. Standard errors are clustered at the school level because this is both the level of randomization and stratification.

3

Results



3.1

Descriptive Statistics

Number of Households Surveyed

Baseline

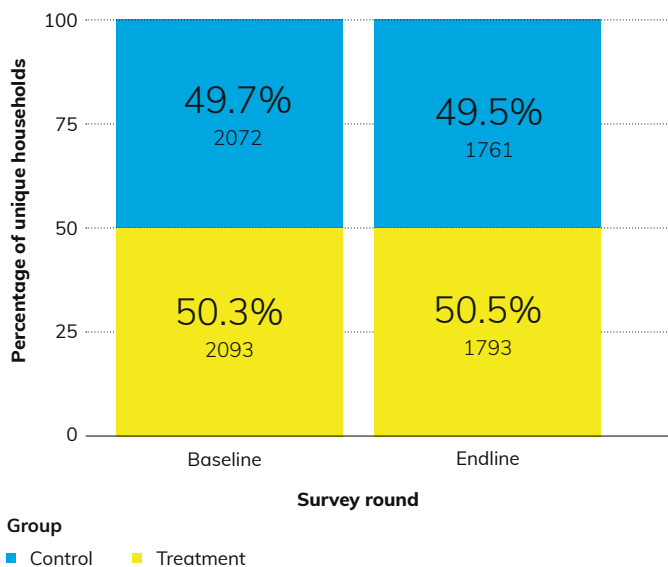
A detailed examination of households surveyed at baseline reveals that we interviewed a total of 4,165 households, 2,093 in the treatment group and 2,072 in the control group. This initial count of surveyed households establishes the sample size, ensuring that both groups are comparable in size for subsequent analysis.

Within these households, the total number of children surveyed was nearly equal, with 2,198 children in the treatment group and 2,175 in the control group. This balance is crucial for the integrity of the study, as it minimizes potential biases that could arise from having significantly different numbers of children in each group.

Endline

At endline, we surveyed a total of 3554 households. Across these households, 1761 children were observed in the control group, while 1793 were observed in the treatment group. We observed a degree of attrition at endline, which could be attributed to various factors such as, parents being unreachable and other logistical challenges.

Figure 3: Proportion of unique households at baseline and endline



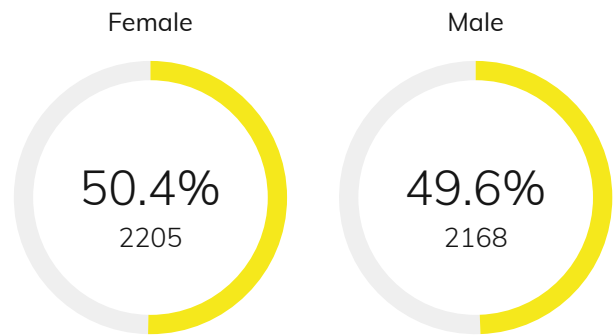
Respondent Gender

At the baseline, the survey included a total of 4,373 children. Of these, 2,205 were female, representing 50.4% of the total sample. The remaining 2,168 respondents were male, accounting for 49.6%. This nearly equal distribution suggests a balanced representation of genders in the initial phase of the study.

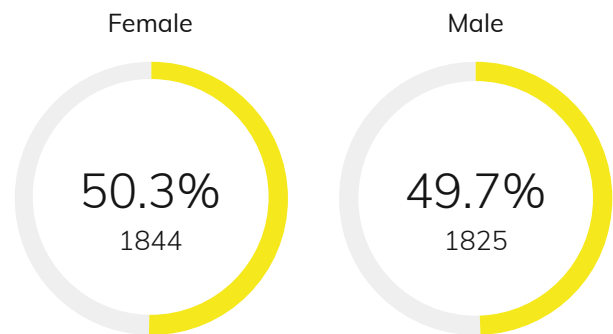
By the endline, the total number of respondents had decreased to 3,669. Of these, 1,844 were female, making up 50.3% of the respondents. Meanwhile, 1,825 respondents were male, representing 49.7% of the total. The gender distribution at the endline remained almost identical to that at the baseline, maintaining a balanced representation of both genders throughout the study.

Figure 4: Gender of respondents at Baseline and Endline

Baseline



Endline

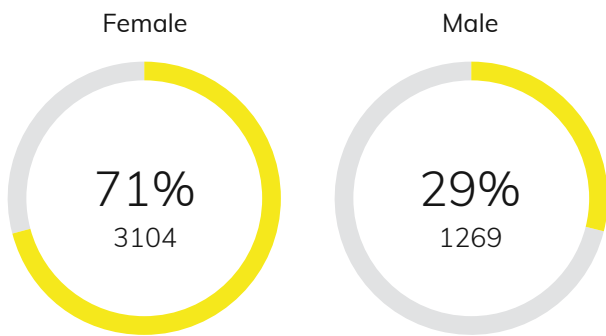


Caregiver Gender

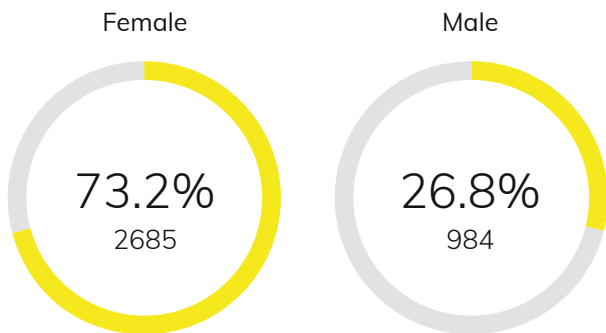
In terms of caregiver gender, we found that the majority of caregivers were female both at the baseline and endline. At baseline, we enrolled 3104 female caregivers, making up 71% of the sample, and 1269 male caregivers, making up 29% of the total sample. By the endline, we were able to track 2685 female caregivers, constituting 73.2% of the total, while 984 caregivers were male, making up 26.8%.

Figure 5: Caregiver gender at Baseline and Endline

Baseline



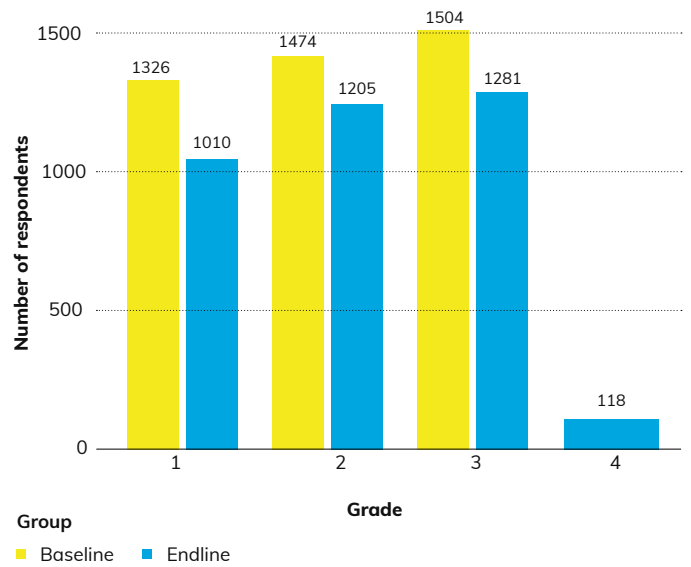
Endline



Grade

Baseline grade levels show that we initially had 1326 students in Grade 1, 1474 in Grade 2, and 1504 in Grade 3. By endline, the numbers shifted with 1010 students in Grade 1, 1205 in Grade 2, 1281 in Grade 3, and 118 in Grade 4. The data indicates fluctuations across grade levels, suggesting potential transitions such as student promotions or attrition, given that a child would not typically be in the same grade after a year. Grade 4 appears exclusively at the endline, indicating progression through successive grades during the period of the intervention.

Figure 6: Grade of respondents at Baseline and Endline



Multiple and Single-Child Households in the sample

As demonstrated in figure 7, our sample had two types of households—those with multiple children who were surveyed from the same house and those with a single child who was surveyed from that house. In baseline, the control group had 201 multi-child households (4.596%) and 1974 single-child households (45.141%), while the treatment group had 200 multi-child households (4.574%) and 1998 single-child households (45.689%). By endline, the control group had decreased to 100 multi-child households (2.726%) and 1711 single-child households (46.634%), whereas the treatment group had 125 multi-child households (3.407%) and 1733 single-child households (47.234%).

Figure 8: Caregiver Education Level at Baseline and Endline

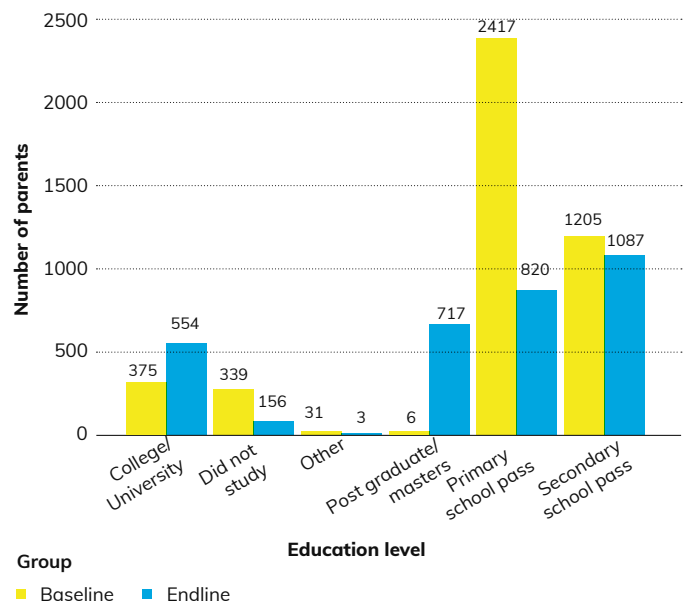
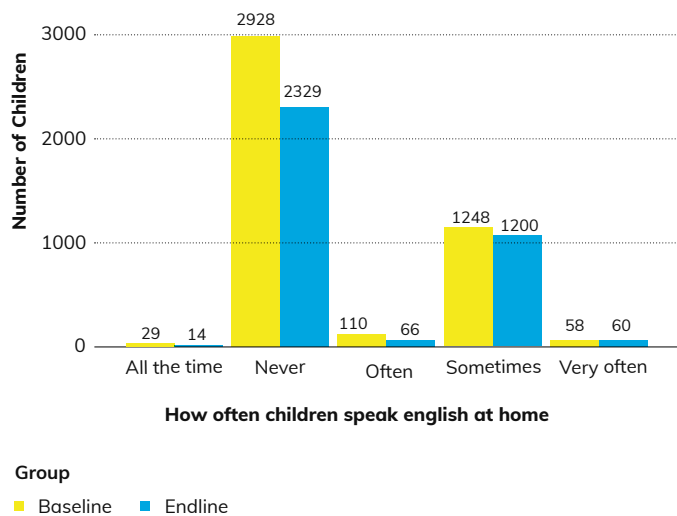


Figure 9: How often children in our sample speak English at home



Caregiver Education Levels

At baseline, the distribution of caregiver educational levels was as follows: 375 individuals had completed college or university, 339 had not pursued any formal education, 31 were categorized under ‘other’ education- ‘other’ in this scenario implies any form of education that we had not listed as one of our choices, 6 had completed postgraduate or master’s degrees, 2417 had passed primary school, and 1205 had passed secondary school.

At endline, the educational attainment levels showed significant changes: 554 individuals had completed college or university, 156 had not pursued any formal education, 3 were categorized under ‘other’ education, 717 had completed postgraduate or master’s degrees, 820 had passed primary school, and 1087 had passed secondary school. It is also worth noting that the data on caregiver education is self-reported, it is therefore subject to bias.

Asset Characteristics

At baseline, 3995 households (91.36%) reported owning a television, while 378 households (8.64%) did not. 2530 households (57.85%) reported having a smartphone, whereas 1843 households (42.15%) did not own a smartphone. It is also important to note that some households had a television at baseline, but it was not functional, and this could have potentially influenced engagement with the intervention.

At endline, out of 3669 households, 3066 (83.57%) owned a functional television, and 603 (16.43%) did not have a functional television. 2460 households (67.05%) reported owning a smartphone, while 1209 households (32.95%) did not own a smartphone. At endline, smartphone ownership decreased to 69% in the treatment group and 65% in the control group, while

Attrition Rates

At endline, the attrition rates were 15.47% for the treatment group and 16.74% for the control group. These rates indicate the proportion of participants who did not complete the study for reasons such as opting out of the study or were not reachable at endline due to factors such as unreachable phone numbers or bad weather conditions. The slightly higher attrition rate in the control group compared to the treatment group showcases differences in participant retention between the control group and treatment group. This could be attributed to factors such as parents in the control group forgetting about the study or being unreachable. However, as demonstrated in the previous section, the study is sufficiently powered to detect meaningful results.

Table 5: Summary of baseline and endline household characteristics

	Treatment	Control	Treatment	Control
Number of surveyed households	2093	2072	1793	1761
Number of kids	2198	2175	1858	1811
Mean age	6.97(2.10)	7.04(2.06)	8.54 (1.28)	8.63 (1.34)
Proportion of female respondents	0.51 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)
Mean grade	2.04 (0.80)	2.04 (0.82)	2.05 (0.80)	2.05 (0.82)
Proportion of multi-child households	0.05 (0.21)	0.05 (0.21)	0.03 (0.18)	0.03 (0.17)

■ Baseline ■ Endline

We analysed various characteristics of our sample, such as gender, age, grade, and other household characteristics. The grade distribution shows that despite a reduction in sample size, 31% of kids were in Grade 1, 34% in Grade 2, and 35% in Grade 3. On average, participants watch TV for 1.23 hours per week, and a significant 87% watch more than 30 minutes of TV daily, with 86% specifically watching Akili Kids. Households typically have about 6 members, with 84% having a TV and 67% having a smartphone. Only 4% of caregivers have primary education or less, indicating a higher level of education among most caregivers.

Additionally, 17% of households speak English, and there is a considerable range in the number of books owned, with an average of approximately 17 books per household. The standard deviations for these variables indicate varying degrees of spread in the sample. For instance, the SD for household size (2.57) and number of books (13.58) suggests a wide range of values, whereas the SD for watching more than 30 minutes of TV (0.34) and watching Akili Kids (0.35) indicates less variability, as illustrated in table 6.

Balance and Attrition

Balance at Baseline

As indicated by Table 7, characteristics across households at baseline represented a favourable balance. The shortfall in sample size did not result in a significant loss of power as the intra-cluster correlations (ICC) for our outcomes are relatively low, ranging from 0.07 for some of the social and emotional learning outcomes to 0.40 for some of the gender outcomes. We find that the study is sufficiently powered to detect a minimum effect size of roughly 0.09 to 0.2 standard deviations, as described in the previous section.

Table 7: Balance Table

Stratification Variables	Treatment	Control	Difference	P-value
Proportion of Female students	0.51	0.50	0.01	0.60
Asset score	4.02	3.96	0.06	0.50
Mean grade	2.04	2.04	0.00	0.96
Caregiver attended college	1.39	1.37	0.02	0.60
TV hours watched per week	0.01	-0.01	0.03	0.66
Baseline Outcomes				
Reading fluency index	3.01	3.03	-0.02	0.72
Comprehension Score	4.90	4.91	-0.01	0.97
Gender stereotype score(mean)	0.60	0.60	0.00	0.60
Gender role score	1.40	1.45	-0.05	0.85
In-group attitude score	2.05	2.00	0.05	0.87
In-group activity score	2.40	2.25	0.16	0.64
Confidence score	9.60	9.52	0.08	0.36
Curiosity score	9.53	9.49	0.04	0.64

At endline, the attrition rates were 15.47% for the treatment group and 16.74% for the control group.

Attrition Rates

At endline, the attrition rates were 15.47% for the treatment group and 16.74% for the control group. These rates indicate the proportion of participants who did not complete the study for reasons such as opting out of the study or were not reachable at endline due to factors such as unreachable phone numbers or bad weather conditions. The slightly higher attrition rate in the control group compared to the treatment group showcases differences in participant retention between the control group and treatment group. This could be attributed to factors such as parents in the control group forgetting about the study or being unreachable. However, as demonstrated in the previous section, the study is sufficiently powered to detect meaningful results.

3.2

Compliance

Defining Compliance

Measuring compliance for a free-to-air broadcast show posed several challenges:

Will the treatment group take up the intervention, and if yes, how do we accurately measure this over a period of 10 months?

Can the control group be preserved by actively controlling advertising the show, and managing how often and when the show is aired?

What is an effective way to gauge the dosage with which the treatment group engaged with the show while accounting for recall bias amongst a very young target audience?

While there is no perfect measure for compliance, the research team attempted to capture this data through a multi-tiered process:

- 1 Self reported**
Have they watched Nuzo and Namia at all, and if yes, how many episodes?
- 2 Character recognition**
Can they identify a character in a show in response to an image prompt?
- 3 Episode specific**
Children were shown clips from each episode to see if they could
 - recall if they watched that episode
 - what happened before/after that clip
 - identify the learning component correctly

Compliance Estimates

Self-reported measure:

The first measure is a self-reported dummy indicator of whether the participant has ever watched a Nuzo and Namia show. By this measure, 79% of the treatment and 31% of the control groups have watched the show. This also includes a self-reported count of the number of Nuzo and Namia episodes the participant has watched. Because the episodes were re-aired, we recoded responses between 13 and 25 as having watched 13 episodes, the number of unique Nuzo and Namia episodes aired. By this measure, the mean number of episodes watched was 3.9 in the treatment group and 1.2 in the control group. We also included episode-by-episode questioning of our instrument to try to gauge compliance at the episode level.

Character recognition:

The next measure of compliance is a question on character recognition where children are shown the image of each of the three main characters and asked to name them. Nuzo and Namia might be easy to infer based on gender and reading the name of the show in the previous question. However, the third main character, who appears in every episode, has a very unique name that is virtually impossible to guess or infer: Bubelang. Responses are coded as correct by enumerators if they correctly identify the name or come close to it. By this measure, 75% of the treatment and 22% of the control were able to identify at least one character in the show. Specifically, 74% and 21% of the treatment and control group respectively were able to identify the character Bubelang.

Episode specific knowledge:

The next measure of compliance aggregates episode-level responses to a participant remembering watching a particular episode after being shown a clip from the given episode. By this measure, respondents in the treatment group watched an average of 4.6 episodes, while respondents in the control group watched an average of 1.4 episodes. Having been shown the episode clip, the participants were asked what happens next in the show to gauge their recall of the episode. By this measure, participants correctly responded 15% of the time in the treatment group

and 5% of the time in the control group. Participants were also asked what the show's characters were learning in the clip to gauge how accurately participants interpreted the intended learning objectives. By this measure, participants responded correctly 27% of the time in the treatment group and 20% of the time in the control group. The differences in compliance between treatment and control groups are statistically significant at the 0.01 level using the Bonferroni correction for multiple hypotheses.

Overview of SMS Engagement data

a. Parent response rates over time

Parents' engagement with the weekly SMS quiz varied greatly, with 2061 parents contacted weekly, 71% female and 29% male. Response rates showed significant ups and downs, hitting a high of 77.10% non-response on March 16, 2024, and a low of 60.31% on July 15, 2023. On average, about 65% to 70% of parents didn't respond, while around 30% to 35% did participate in the quiz.

We did not observe a clear pattern in how responses changed over time. Some days, more parents joined and engaged with the quiz, while others had fewer parents engaging with the quiz. On average, the overall response rate to the quizzes was 31%, with the highest at 41.48% and the lowest at 22.90% in the last week of encouragement data collection. It's important to note that no money or rewards were given to encourage parents to participate.

b. Proportion of correct responses over time

In this section, we look at how many responses to the question, "Which country did Nuzo and Namia visit in this episode?" were correct. The distribution of correct and incorrect responses throughout the intervention period reflects that, on average, approximately 63% of valid responses were correct and 37% incorrect. The highest proportion of correct responses was observed on October 28th, 2023, at 86.18%, while the lowest was on August 5, 2023, with only 28.35% correct. Generally, correct responses fluctuated between 55% and 80% across different weeks, reflecting varying levels of accuracy in parent submissions. Conversely, incorrect responses ranged from 20% to 45%.

c. Impact of parent characteristics on response rates

Continuous monitoring of engagement data prompted the team to analyse how various parent characteristics impact their engagement with the SMS quiz; an overview of this analysis is demonstrated in table 9. Gender differences reveal that females have a slightly higher probability of responding to the SMS quiz compared to males and providing at least one valid response. Moreover, females tend to respond more frequently and also have a higher number of valid responses. This indicates that women were more engaged and accurate in their participation.



Educational attainment emerges as a significant predictor of engagement and response validity. Parents with some primary education show a higher likelihood of responding and providing at least one valid response, although their total and valid responses were not statistically significant. Those with a primary pass exhibit a higher probability of responding and valid responses, with significant increases in the number of responses and valid responses. Secondary and higher education levels further amplify these effects, with college/university graduates and postgraduate parents showing the highest engagement and accuracy, reflecting a strong correlation between higher education and better SMS quiz participation.

Asset ownership showed mixed results. TV ownership did not significantly impact the probability of responding or the number of responses. However, smartphone ownership slightly decreases the probability of providing at least one valid response, though it does not significantly affect other measures. Non-smartphone ownership and electricity access have positive but not statistically significant effects on the likelihood of responding and providing valid responses, except for electricity, which increases the number of responses and valid responses.

Regional differences highlight varied impacts on SMS quiz engagement. Counties like Kakamega and Kisumu did not show significant differences in response rates compared to the reference county, Kajiado. However, Kiambu and Meru counties exhibited significantly lower response rates, with Kiambu showing a reduction in the number of responses and valid responses and Meru displaying a similar trend with lower response counts and valid responses. These regional disparities suggest that location plays a role in influencing parent engagement with the SMS quiz, potentially due to varying access to resources or differing local conditions.

Effects of Encouragement on Compliance

The encouragement design led to a notable increase in self-reported compliance. Column (1) in table 10 indicates that the likelihood of a child reporting they had watched the show at least once rose by 46.2 percentage points (p -value <0.001), compared to a baseline of 31.1% in the control group—a 149% increase. Column (3) shows that consistent watching (defined as watching six or more episodes) also increased by 21.3 percentage points (p -value <0.001), a 345% rise from the control group's mean of 6.18%. These estimates remained consistent even after controlling for school-level factors, indicating that the

encouragement design effectively boosted both occasional and consistent viewing. However, self-reported measures may be biased upward due to experimenter demand effects.

To mitigate this, columns (5) to (10) of Table 4 examine recall-based compliance measures, which are less susceptible to demand effects. On average, 10.7% of control children could correctly identify the three main characters, while treatment students were 35.6 percentage points more likely to recall all three names (p -value <0.001). Accurately recalling specific episode content was more challenging; only 3.5% of control children identified six or more episodes, while the encouragement increased this by 8.9 percentage points (p -value <0.001). Additionally, recall of specific episodes improved by 10 percentage points (p -value <0.001).

Across all compliance measures, the encouragement treatment significantly increased show watching. Recall-based measures indicate overall lower viewing in the control group—about a third as high for character identification and half as high for consistent watching, compared to self-reports. Treatment effects were smaller but less influenced by demand effects on recall measures, which may be tied to factors like recall ability. Table 5 shows evidence of heterogeneity in recall-based effects by baseline literacy; children with higher initial literacy were 6.8 percentage points more likely to identify characters (p -value=0.023) and showed a larger increase in consistent watching, both on self-reported and recall-based measures, compared to children with lower literacy levels.

Table 8: Effects of Encouragement on Compliance

Self-Reported	Ever Watch		Consistent Watch	
	1	2	3	4
Treatment	0.462*** (0.019)	0.462*** (0.019)	0.213*** (0.015)	0.213*** (0.015)
School Controls	No	Yes	No	Yes
Mean Dep. Variable	0.311	0.311	0.0618	0.0618
(R ²)	0.284	0.286	0.136	0.136
Observations	3,669	3,669	3,669	3,669

Recall	Identify Characters		Consistent Watch		% Episodes Identified	
	5	6	7	8	9	10
Treatment	0.356*** (0.020)	0.356*** (0.019)	0.089*** (0.014)	0.087*** (0.013)	0.100*** (0.010)	0.099*** (0.010)
School Controls	No	Yes	No	Yes	No	Yes
Mean Dep. Variable	0.107	0.107	0.0353	0.0353	0.0462	0.0462
(R ²)	0.217	0.221	0.082	0.085	0.143	0.148
Observations	3,669	3,669	3,669	3,669	3,669	3,669

Note

Each column presents estimates corresponding to an OLS regression for a different outcome or specification. All estimates include quadruple fixed effects. Even numbered columns include school-level controls: share of boys, asset indicator, average grade, average reading score, school size. Cluster-robust standard errors at the school level in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.1

3.3

Treatment effects on main outcomes

Descriptive Statistics

Table 11 provides an overview of how each outcome performed through a descriptive snapshot. The Literacy Index is a cumulative score of two components: oral fluency and comprehension. We find that on average, children in the study obtained a score of 0.02 on literacy. The Gender Index is a composite index of the 4 gender sub-items: gender stereotype knowledge, gender role behaviors, in-group attitudes, and activities. A negative gender index score indicates biased gender attitudes whereas a positive index score indicates progressive or neutral gender attitudes. We find the Gender Index mean to be 0 at endline. The SEL score is an average of the two SEL sub-items: confidence and curiosity. Scores were generally high for this outcome measure and we observe an average of 18.82 on the SEL score.

Table 10: Effects of Encouragement on Compliance

	Mean	Standard deviation	N
Literacy Index	0.02	0.75	3577
Reading Fluency Index	0.01	0.78	3577
Comprehension Score	4.59	3.37	3669
Gender Index	0.00	0.11	3662
Gender Stereotype score(mean)	0.56	0.18	3669
Gender Stereotype score(sum)	7.85	2.47	3669
Gender Role score	1.69	4.45	3669
In-group attitude score gender	1.99	3.53	3669
In-group activity score gender	2.41	4.38	3669
SEL score	18.82	3.68	3590
Confidence score	9.41	2.16	3632
Curiosity score	9.39	2.14	3620

Intent to Treat Estimates on student learning outcomes

Note: Each column represents estimates for an OLS regression corresponding to a different outcome or specification. All estimates include controls for the baseline dependent variable and quadruplet fixed effects. Cluster robust standard errors at the school level in parenthesis. Significance ***p<0.01, **p<0.05, *p<0.1

Overview of Results

Table 11 summarises the ITT estimates as observed in the study. We found positive and slightly improved measures on SEL across all stratification methods. The Literacy Index and Gender Index display positive effects; however, these differences between treatment and control groups do not reach statistical significance. Let us examine Panel A, which summarises the results by the primary outcome measures of interest. We have followed three stratification approaches: the first accounts for the quadruple fixed effects at the school level owing to uniform clusters during sampling; the second controls for baseline levels of learning as indicated by the outcome variables; the third accounts for stratification variables at the school level such as proportion of female students at school parental education, etc.

We find that when we only control for the quadruple fixed effects and baseline learning levels, we see a statistically significant impact on the gender index. When all three stratifications are accounted for, we see a statistically significant impact on the gender index.

The more interesting insights are obtained from Panels B and C, which examine the impact on the sub-items based on the domain of learning. We do not find significant impacts on cognitive learning outcomes such as oral fluency or comprehension. However, what is incredibly interesting is the impact on two sub-items under the non-cognitive learning domain: gender role and curiosity. The gender role sub-item looks into the behaviors of the children. Specifically, it asks the question: do children exhibit positive gender behaviors through their actions as a result of watching Nuzo and Namia? We find that they do. This implies that the show has been able to positively influence actual behaviors towards a more progressive gender outlook. We also observe a positive impact on curiosity towards reading amongst children as a result of watching the show. This affirms our hypothesis that mimicking positive behaviors over time and cultivating a curiosity for reading might lead to gains in literacy over time.

Table 11: Panel A: Effects on Primary Outcomes

Variable	Literacy index			Gender Index			SEL Index		
	1	2	3	4	5	6	7	8	9
Treatment	0.043 (0.056)	0.051 (0.053)	0.043 (0.049)	0.052 (0.045)	0.056 (0.045)	0.053 (0.046)	0.086* (0.046)	0.079* (0.046)	0.078* (0.045)
School Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Individual Controls	No	No	Yes	No	No	Yes	No	No	Yes
(R ²)	0.338	0.349	0.386	0.185	0.188	0.200	0.083	0.087	0.091
Observations	3,577	3,577	3,577	3,669	3,669	3,669	3,571	3,571	3,571

Table 12: Intent to treat estimates on secondary outcomes

Panel B: Effects on Literacy					
Variable	Treatment	SE	Mean DV	(R ²)	N
Non-word reading (z)	-0.003	(0.0046)	0.001	0.387	3,617
Letter Recognition (z)	-0.006	(0.052)	0.001	0.210	3,630
Oral fluency (z)	0.035	(0.037)	0.003	0.428	3,638
Reading Comprehension (z)	0.080*	(0.048)	-0.000	0.318	3,669
Listening Comprehension Score (z)	0.084	(0.056)	0.000	0.293	3,669
Panel C: Effects on Gender					
Gender Stereotype (z)	-0.036	(0.042)	0.000	0.080	3,669
Gender Consistency Behavior score (z)	0.111***	(0.039)	-0.000	0.247	3,669
Gender Consistency Trait score (z)	0.029	(0.026)	-0.000	0.185	3,669
Gender Bias. Willingness to Interact (z)	0.059	(0.066)	-0.000	0.405	3,669
Ingroup Willingness to Interact (z)	-0.011	(0.037)	-0.000	0.383	3,669
Outgroup Willingness to Interact (z)	-0.070	(0.049)	-0.000	0.190	3,669
Gender Bias Interactions (z)	0.028	(0.066)	-0.000	0.305	3,669
Ingroup Interactions (z)	0.039	(0.041)	-0.000	0.247	3,669
Outgroup Interaction (z)	0.006	(0.047)	-0.000	0.153	3,669
Panel D: SEL					
Confidence Score (z)	0.028	(0.042)	0.004	0.075	3,548
Curiosity score (z)	0.109**	(0.045)	-0.005	0.075	3,547

Note: Each row presents estimates for an OLS regression corresponding to a different outcome. Each panel presents components of primary indices for our three main outcomes. All estimates include controls for the baseline dependent variable, school controls and quadruple fixed effects. Schools controls: share of boys, asset indicators, average grade, average reading score, school size. Cluster robust standard errors at the school level in parenthesis. Significance. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3.4

Local Average Treatment Effects (LATE/CACE)

We estimated the local average treatment effects (LATE) on compliers in our study by adjusting the main analytical model to replace the treatment indicator with a compliance measure (e.g., ever watched the show, consistent watching). However, simply using compliance measures can lead to biased estimates, as watching the show may be associated with other factors influencing outcomes. For instance, higher SES households might both watch more educational TV and perform better on literacy tests. To address this, we leveraged our randomized encouragement design, which nudges treated households to watch the show, as an instrumental variable for compliance.

Table 13 presents the 2SLS estimates for our three main indices across different panels, with each column reflecting a different compliance measure. Each compliance measure captures distinct aspects of engagement, each with unique limitations; thus, considering a range of measures helps to bound the effects. Consistent with ITT results, literacy and gender estimates lack precision. For literacy, point estimates range from 9.4% of a standard deviation with self-reported “ever watching,” to 48.2% with recall-based measures of consistent watching (i.e., correctly identifying six or more episodes), though the 95% confidence interval here is wide, from -70.5% to 166.8%. Gender index estimates are similarly varied.

For socio-emotional learning (SEL), results show statistically significant effects at the 10% level. Self-reported “ever watching” suggests an 18.6% increase in SEL ($p=0.068$), while recall-based “ever watching” using character identification shows a 40.7% increase ($p=0.067$). These results reflect smaller first-stage treatment effects for recall-based compliance measures.

...higher SES households might both watch more educational TV and perform better on literacy tests.

Table 13: 2SLS estimates

Compliance Measure	1	2	3	4	5
	Ever Watch	Character ID	Consent Watch	CW (Correct)	% episodes
Panel A Literacy Index Compliance Measure	0.094 (0.120)	0.202 (0.257)	0.121 (0.154)	0.482 (0.603)	0.433 (0.542)
Observations	3,577	3,577	3,577	3,577	3,577
(R ²)	0.244	0.242	0.250	0.254	0.261
First Stage F Test Statistic	1032	313.2	677.3	97.03	272.9
Panel B Gender Index Compliance Measure	0.112 (0.098)	0.243 (0.213)	0.145 (0.127)	0.583 (0.507)	0.518 (0.447)
Observations	3,669	3,669	3,669	3,669	3,669
(R ²)	0.106	0.101	0.106	0.091	0.106
First Stage F Test Statistic	1044	314.6	677.3	97.01	275.9
Panel C SEL Index Compliance Measure	0.186* (0.101)	0.407* (0.222)	0.242* (0.131)	0.983* (0.554)	0.860* (0.470)
Observations	3,571	3,571	3,571	3,571	3,571
(R ²)	0.015	0.007	0.018	-0.036	0.007
First Stage F Test Statistic	1007	296.8	644	90.31	264.6

Note: 2SLS estimates where the compliance measure (in columns) are the main endogenous variables, instrumented by the treatments indicator. All regressions include quadruple fixed effects and the baseline outcome as control. Kleibergen-Paap first stage F-statistics presented. Cluster robust standard errors at the school level in parenthesis. Significance: ***p<0.01, **p<0.05, *p<0.1

3.5

Heterogeneity

The encouragement to watch Nuzo & Namia may have impacted sub-groups differently due to several factors. Since the show aired in English, language barriers could have influenced engagement, while socio-economic factors may also have contributed. Table 13 displays intent-to-treat (ITT) effects across various sub-groups and includes p-values to assess differences in coefficients between groups. To strengthen statistical power, we incorporated school characteristics and individual controls in our calculation model. Given the multiple sub-group analyses, these findings should be considered suggestive.

Household language appears as an important factor, especially as Nuzo & Namia aired in English, while about 75% of our sample primarily spoke Swahili at home. Panel A of Table 13 highlights this variation, showing that children from non-English-speaking households experienced economically small and statistically insignificant gains across the three main indices (point estimates ranged from 1.6% to 4.4% of a standard deviation for literacy and SEL). In contrast, children from English-speaking households saw greater effects, with a 15.2% increase on the literacy index ($p=0.081$) and a 20.4% increase on the SEL index ($p=0.021$). However, we do not have sufficient statistical power to confirm a significant difference in literacy effects between English- and non-English-speaking homes ($p=0.110$).

Socio-economic status may also play a role in the intervention's effects. For example, parents with higher education levels could reinforce concepts or invest in additional educational resources that complement the show. We examine this in Panels C and D of Table 13, using proxies for parental education (completion of secondary school or higher) and the presence of books in the household. Findings suggest that children of less-educated parents had a significant 9.7% increase in their Gender Index score, a difference that is statistically significant compared to children of more educated parents ($p=0.002$). However, literacy gains did not vary significantly by parental education or household book availability, indicating that book access was not a major factor.

Further analyses (Panels B, E, and F) looked at dimensions such as gender, household size, and baseline outcomes, finding no significant heterogeneity in treatment effects across these factors.

3.6

Mechanism

In section 1.2 we outlined the Theory of Change, detailing the mechanisms that could mediate the effects of our treatment on various outcomes. We have categorized these outcomes into intermediate behaviors and underlying mechanisms (preferences, information, time use, and social interactions) and present the findings below.

3.6.1 Intermediate Behaviors

Intermediate behaviors are actions often needed to drive changes in primary outcomes. For instance, a positive shift in reading habits is generally essential for enhancing reading skills, as children are unlikely to improve without a substantial increase in reading behavior. We begin by assessing whether treatment led to any changes in reading behavior.

Our analysis does not reveal any statistically significant effects on the measured intermediate behaviors. This lack of built habits may partly explain the absence of a significant impact on overall literacy outcomes in this study.

3.6.2 Preferences

Overall, children in the treatment group generally did not identify strongly with the show's main characters—Nuzo, Namia, and Bubelang—with over 50% of them indicating that each character was “not at all like me.” To assess whether relating to characters impacts learning, we conducted a mediation analysis within the treatment group.

Although this analysis is correlational (as viewing intensity wasn't randomly assigned), regressing outcomes on the compliance measure “Consistent watch” for children who viewed at least one episode suggests a possible dosage effect on literacy. However, a mediation analysis testing if character relatability (i.e., seeing oneself in the character most aligned with literacy skills) influenced this effect shows no mediation impact (average causal mediation effect (ACME) near zero across all outcomes).

For gender and socio-emotional learning (SEL) outcomes, we find no statistically significant dosage effect, making it difficult to determine if character relatability plays any role here, with ACME estimates again close to zero. Additionally, regarding reading preferences—interest, enjoyment, or attitudes toward challenging books—no significant treatment effect was found.

Table 14: ITT Effects-Mechanisms

Mechanism	Intermediate Behavior			Preferences	Time Use	Indirect Mechanism	
	1	2	3				4
Dep. Variable	Read for enjoyment	Parent encouragement	Ask for Books	Parent Read Stories	Positive Affect	TV Hours	Social Interaction
Treatment	0.043 (0.039)	0.004 (0.041)	-0.006 (0.031)	-0.016 (0.035)	-0.121 (0.493)	0.013 (0.065)	0.003 (0.054)
School Controls	No	No	No	No	No	No	No
Individual Controls	No	No	No	No	No	No	No
0.019	0.019	0.019	0.020	0.009	0.052	0.147	0.024
Observations	3,669	3,669	3,669	3,669	3,662	3,669	3,654

Note: We column presents estimates for an OLS regression corresponding to a different outcome or the specification. All estimates include quadruple fixed effects. Cluster-robust standard errors at the school level in parenthesis. P-value FDR adjusted for multiple hypothesis testing. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 15: Correlational Mediation Analysis

Dep. Variable	Literacy Index	Gender Index	SEL Index
Mediation Effect of character alike	-0.000071 (0.98)	-0.00007 (0.77)	0.0007 (0.73)
Direct Effect of consistent watch	0.175*** (0.00)	-0.004 (0.52)	0.062 (0.14)
School Controls	No	No	No
Individual Controls	No	No	No
Observations	1,431	1,464	1,411

Note: Sample subset to treatment group children who reported to have at least watched one episode. Interpretation of mediation and direct effects is not causal. P-values calculated from bootstrapped confidence intervals in square brackets.

3.6.3 Information/Skills

Unlike other educational TV interventions delivering specific information, this show models strategies for reading and comprehension. According to the theory of change in section 1.2, this effect may partially work through increased reading. By examining items in the literacy tests aligned with show content (e.g., recall, scanning, inference, and prediction), we assess how much these strategies contribute to the results.

While the overall effect on our literacy index isn't significant, children did show improvement in reading comprehension—a skill the show targeted. Moreover, children in the treatment group generally recalled show content, suggesting that Nuzo & Namia successfully conveys information that sticks.

We initially hypothesized that children with the lowest baseline reading scores would benefit the most, as they likely have the largest knowledge gaps (La Ferrara, 2016). However, this pattern wasn't observed. Given that children primarily learned higher-level skills like comprehension, it's possible that these skills were challenging for those with lower initial reading abilities.

3.6.4 Time Use, Social Interaction and Barriers

We find no evidence of treatment effects on children's time use or their social interactions with peers and parents. Given that children did not seem to request more books, a lack of books at home could be a barrier to developing reading habits. Descriptively, this appears to be an issue, as 44% of children report having “none or very few, like 0-10 books” at home.

Our heterogeneity analysis does not show that the baseline number of books significantly affects outcomes; however, it remains uncertain whether a general scarcity of books may pose a barrier for all children in the sample, despite some variation within the group.

44% of children report having “none or very few, like 0-10 books” at home

Given that children did not seem to request more books, a lack of books at home could be a barrier to developing reading habits.

According to GeoPoll data collected in July and December 2022, Akili Kids reaches an average of 1,238,000 children during Ubongo's program times

3.7

Cost Effectiveness Analysis

The CEA followed the ingredient method discussed in **McEwan (2012) and Dhaliwal et al. (2013)**. Closely collaborating with Ubongo, we have categorised different costs incurred as a result of producing, airing, and marketing the show. Since the show was produced as part of this experiment, we extrapolate various costs from Ubongo's other productions as a proxy measure for the cost of the program outside the context of this evaluation.

Specifically, we compiled over 5,600 individual cost entries from Ubongo. Using task descriptions from an internal project management system in conjunction with financial data, we organized all production and implementation costs into 18 distinct categories, triangulating the categorization process between two researchers.

We also estimated the scaled costs for encouragement (e.g., parent outreach and SMS reminders), beneficiary expenses (e.g., electricity for TV usage), and distribution fees (e.g., channel airing costs). "At scale" here refers to the per-child cost in each category, multiplied by the total number of children reached by the Kenyan TV channel broadcasting Nuzo & Namia. According to GeoPoll data collected in July and December 2022, Akili Kids reaches an average of 1,238,000 children during Ubongo's program times. We derived these costs using internal records, Kenya Power and Lighting Company's average monthly domestic tariff rates, and the Akili Kids rate card for broadcast clients.

Since the intervention involved a TV program, we categorized production of the show as an investment cost. Using U.S. and U.K. Generally Accepted Accounting Principles (GAAP) and similar shows like Sesame Street, Dora the Explorer, and Ubongo Kids, we estimate that each season remains relevant and broadcastable for around ten years. Thus, we allocate one-tenth of the investment cost annually in our cost-effectiveness analysis. Annual costs also include Ubongo's operational overheads, encouragement and targeting efforts, airing fees, and beneficiary expenses. To enable comparisons with previous studies, we report results in real dollars for both the analysis year and 2011.

Note: Sample subset to treatment group children who reported to have at least watched one episode. Interpretation of mediation and direct effects is not causal. P-values calculated from bootstrapped confidence intervals in square brackets.

In comparison to a structured pedagogy program costs which would cost an average of \$8.62 (2020 USD) per child, and a Teaching at the Right Level (TaRL) intervention which would average a costs of \$19.53 or \$27.18 (2020 USD, in the Kenya/LMC context) without and with technology, respectively (Angrist et al., 2023). The estimated cost of this edutainment intervention, at \$1.67 per child (2024 USD), equivalent to \$1.41 in 2020 USD, highlights a promising cost profile.

Table 16: Primary Cost Effectiveness Results

	Literacy	Gender	SEL	Average	Combined
Annual Costs (2024 real USD)	----	----	2064819.01	----	----
Annual Costs (2011 real USD)	----	----	1508454.64	----	----
Total Benefits Accrued (Std Dev.)	53234	64376	102754	734554.67	220364
Benefits Per Child (Std Dev.)	0.043	0.052	0.083	0.059	0.178
Cost Per Child (2024 real USD)	----	----	1.67	----	----
Cost Per Child (2011 real USD)	----	----	1.22	----	----
Cost Per Std. Dev (2024 real USD)	38.79	32.07	20.09	28.11	9.37
Cost Per Child (2011 real USD)	28.34	23.34	14.68	20.54	6.85

Table 17: CEA Results for English Speaking Households

	Literacy	Gender	SEL	Average	Combined
Annual Costs (2024 real USD)	----	----	459673.83	----	----
Annual Costs (2011 real USD)	----	----	335814.96	----	----
Total Benefits Accrued (Std Dev.)	37261.57	31672.34	45127.9	38020.6	114061.81
Benefits Per Child (Std Dev.)	0.18	0.153	0.218	0.184	0.551
Cost Per Child (2024 real USD)	----	----	2.22	----	----
Cost Per Child (2011 real USD)	----	----	1.62	----	----
Cost Per Std. Dev (2024 real USD)	12.34	14.51	10.19	12.09	4.03
Cost Per Child (2011 real USD)	9.01	10.6	7.44	8.83	2.94

Note: The at scale audience in the calculations was reduced proportionally to 207,009, using the share of English speaking households in our endline survey.

4

Learnings and Recommendations



This study uses a large RCT to examine the impact of EdTech including educational TV on learning outcomes. The ITT estimates demonstrate that the Nuzo and Namia TV can be effective in enhancing non-cognitive domains of learning, such as gender attitudes and social emotional development, and may also positively influence literacy outcomes, provided language barriers do not impede progress.

This study provides valuable insight into the practical considerations that need to be taken into account while designing supplemental learning programs through last-mile distribution methods. The results indicate that an intervention such as Nuzo and Namia can be very effective in improving non-cognitive learning domains such as gender attitudes and socio-emotional learning. Hence, government investment in EdTech infrastructure including educational television is likely to promote non-cognitive learning.

However, this study brings to light the difficulty of promoting learning in cognitive domains such as literacy, an outcome that has eluded many learning institutions and continues to be a challenging skill to develop. However, learning often happens in inter-connected systems and it would be interesting to explore how shifts in attitudes and behaviors affect literacy in the long run. This raises the question of the types of learning systems that should come under consideration for improved literacy.

While this study provides important insights on the impact of educational TV, it suffers from several limitations. First, the sample was drawn from public schools, however, there was no formalised partnership with the Ministry of Education (MoE) to allow teachers to use the Nuzo and Namia show as a learning resource and a complement to the traditional teaching resources. Further avenues of research include investigating the impact of educational TV with interactive child-adult co-viewing including caregivers and teachers. Additionally, further insights might be uncovered by investigating the impact of on-demand access of educational TV on learning outcomes. By extension, this may additionally provide insightful knowledge on whether on-demand access to the show on multiple platforms enhances viewership and thus learning outcomes.

The first important system is a formal schooling environment. In this case, our sample was drawn entirely from the public school system in Kenya, thus rendering the Ministry of Education (MoE) an important stakeholder. A preliminary line of inquiry in this research, which was embargoed owing to budgetary constraints and the lack of requisite permissions to work with public schools, was the impact of educational television on teacher effectiveness. We wondered if teachers could use the Nuzo and Namia show as a learning resource to complement their traditional teaching practices. An important lesson learned by the research team is that the government should be included in the research process from the very beginning. Just as we incorporated heavy collaboration from our intervention partner Ubongo, we should have made concerted efforts to include the MoE from the start, including the research design process. This compounded with a change in the MoE's protocols on how to engage with external researchers resulted in them rescinding our permission to access public schools. As a result of failing to engage the MoE early enough, we were unable to examine how the Nuzo and Namia show could effectively supplement the learning that happens in formal schooling environments. We have learned that the government does not merely act as a conduit of access, but as an



important stakeholder in understanding the tools used, the skills we were measuring for, and the overall curriculum design of the intervention.

The second important learning is context and in this case, language. We selected English literacy as the language of measurement and programming for the show because our study sample would be integrated into a school system where English is the official language of instruction. We believed that in a context where most families speak their mother tongue at home, it would be difficult to pick one standard language for the evaluation and decided it might be best to inquire how the show improves school readiness and integration among the study sample. A learning from this is to focus more strongly on decontextualized language skills and operate in a language-agnostic space in future evaluations. The reason for this is that such programs might be more beneficial in impacting decontextualized skills in literacy that overlap across languages, such as inference and prediction. Oral fluency is language-specific (rules of phonics in English are not the same as that in Kiswahili) and may be best taught through mediators such as a teacher. However, the ability to make meaning of text or vocabulary can be built well through such programs that rely less on text-on-screen and more on the child's ability to make meaning.

The third is the sheer scale and variety of considerations that an implementing organisation such as Ubongo must take into account to deliver this content. It is no small feat to design a program that spans over 5 different cultures, 3 languages, and is apt for a neurodivergent audience. Such programming requires time, multiple testing and iterations, and a significant technological investment, among other resources. Ubongo has made considerable efforts in this direction but a big learning for us all was that it takes time to launch something this ambitious. We hope to use the findings from this evaluation to further develop subsequent seasons of Nuzo and Namia, an effort that is already underway from the Ubongo team. It is important to note that data from this study from the very beginning was used in an iterative way to inform programming. This aspect is reflected in how the curriculum in Nuzo and Namia was developed based on pilot data, as well as learnings from the qualitative study to understand how children engage with the show. A cornerstone of

research-practice partnerships is to ensure that the efforts of both teams are complementary to each other and this is reflected well in this study. Both the Busara and Ubongo teams have met regularly over the three-year study design and implementation period to constantly provide feedback to each other on how our work could be improved. Best practices from this initiative include:

Using data from the research team regularly to pivot and improve programming.
Ensuring meaningful continuous data is collected over the lifetime of the project that describes mechanisms that are useful in understanding the impact on the final outcomes.
Partnering with specialists from the Ubongo team to train enumerators and inform field practices which proved to be incredibly important while engaging with children younger than 9 years of age.
And most importantly, creating open channels of communication.

This study reinforces our hypothesis that supplementary educational programs that introduce learning in a fun way to children are beneficial, especially for communities on the margin. We hope to use the learnings from this study to further benefit holistic learning ecosystems and leverage the power of educational technology to create better readers and most importantly, confident and curious humans who believe in and further the cause of gender equality.

References

- Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics* 134(3), 1121–1162.
- 31
- Alan, S., S. Ertac, and I. Mumcu (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics* 100 (5), 876–890.
- Araya, R., E. Arias Ortiz, N. L. Bottan, and J. Cristia (2019). Does gamification in education work? experimental evidence from Chile.
- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics* 60(1), 47–57.
- Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics* 47(2), 1148–1178.
- Bandura, A. and R. H. Walters (1977). *Social learning theory*, Volume 1. Englewood cliffs Prentice Hall.
- Banerjee, A., E. La Ferrara, and V. H. Orosco-Olvera (2019). The entertaining way to behavioral change: Fighting HIV with mtv. Technical report, National Bureau of Economic Research.
- Borzekowski, D., D. Singpurwalla, D. Mehrotra, and D. Howard (2019). The impact of galli galli sim sim on Indian preschoolers. *Journal of Applied Developmental Psychology* 64, 101054.
- Borzekowski, D. L. (2018). A quasi-experiment examining the impact of educational cartoons on Tanzanian children. *Journal of Applied Developmental Psychology* 54, 53–59.
- Borzekowski, D. L., Mondestin, T., & St-Onge Ahmad, S. (2024). “Li ka fe’l tou”: The influence of an educational television program on Haitian children’s gender beliefs. *International Journal of Child, Youth and Family Studies*, 14(4), 125-142.
- Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics* 1(4), 200–232.
- Center, P. R. (2018). Social media use continues to rise in developing countries but plateaus across developed ones.
- Cherewick, M., S. Lebu, C. Su, L. Richards, P. F. Njau, and R. E. Dahl (2021). Promoting gender equity in very young adolescents: targeting a window of opportunity for social emotional learning and identity development. *BMC Public Health* 21 (2299).
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, February). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21 (1), C1&C68.
- Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India.
- Chong, A. and E. L. Ferrara (2009). Television and divorce: Evidence from Brazilian novelas. *Journal of the European Economic Association* 7 (2-3), 458–468.
- Deole, S. S. and T. Zeydanli (2021). Does education predict gender role attitudes?: Evidence from European datasets. GLO Discussion Paper 793 [rev.], Global Labor Organization (GLO), Essen.
- Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. *Education Policy in Developing Countries*.
- Ferrara, E. L., A. Chong, and S. Duryea (2012). Soap operas and fertility: Evidence from Brazil. *American Economic Journal: Applied Economics* 4 (4), 1–31.
- Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75, 259–276.
- Gustafsson-Wright, E., I. Boggild-Jones, and S. Gardiner (2017, September 5). Sect: The standardized early childhood development costing tool.
- Gustafsson-Wright, E. and D. Lee (2021, December 23). Getting both costs and effectiveness right to improve decisionmaking in education.
- International Institute for Educational Planning (IIEP) - UNESCO (2024). Learning and teaching materials.

- Jukes, M. C. H., N. L. Mgonda, J. J. Tibenda, P. Gabrieli, G. Jeremiah, K. L. Betts, J. Williams, and K. L. Bub (2021). Building an assessment of community-defined social- emotional competencies from the ground up in tanzania. *Child Development* 92, 1095–1109.
- Kearney, M. S. and P. B. Levine (2015). Media influences on social outcomes: The impact of mtv’s 16 and pregnant on teen childbearing. *American Economic Review* 105(12), 3597–3632.
- Kearney, M. S. and P. B. Levine (2019). Early childhood education by television: Lessons from sesame street. *American Economic Journal: Applied Economics* 11 (1), 318–350.
- Kenya Ministry of Education (2019). Basic education statistical booklet. Technical report, Ministry of Education.
- La Ferrara, E. (2016). Mass media and social change: Can we use television to fight poverty? *Journal of the European Economic Association* 14 (4), 791–827.
- Mares, M.-L. and Z. Pan (2013). Effects of sesame street: A meta-analysis of children’s learning in 15 countries. *Journal of Applied Developmental Psychology* 34 (3), 140–151.
- McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness* 4 (2), 189–213.
- McKenzie, D. (2022, April 19). Why i am now more cautious about using or recommending matched pair randomization and like matched quadruplets instead.
- Merga, M. K. and S. M. Roni (2017). The influence of access to ereaders, computers and mobile phones on children’s book reading frequency. *Computers & Education* 109, 187–196.
- Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review* 109(4), 1426–1460.
- Murunga, J. (2021). Radio, tv and internet audience statistics in kenya â q1 2021.
- Perski, O., A. Blandford, C. Garnett, D. Crane, R. West, and S. Michie (2020). A self-report measure of engagement with digital behavior change interventions (dbcis): development and psychometric evaluation of the âdbci engagement scaleâ. *Translational Behavioral Medicine* 10(1), 267–277.
- Powell, D. (2013). A new framework for estimation of quantile treatment effects nonseparable disturbance in the presence of covariates.
- Tadesse, S. and W. Muluye (2020). The impact of covid-19 pandemic on education system in developing countries: A review. *Open Journal of Social Sciences* 8, 159–170.
- UNESCO (2022). Persistent teacher gap in sub-saharan africa jeopardizing education recovery.
- UNESCO (2023). Literacy rate, adult male.
- USAID (2016). Early grade reading assessment (EGRA) toolkit. Technical report. Usawa (2021). Uwezo learning assessment. Technical report, Usawa Agenda.
- Wawire, B., F. Kiroro, N. Muhia, C. Asego, R. Perakis, and G. Smarrelli (2023). From evidence to policy-making: What should be in kenyaâs next education and training gender policy?
- WorldBank, UNESCO, UNICEF, USAID, FCDO, and B. . M. G. Foundation (2022). The state of global learning poverty: 2022 update.
- Yu, L., J. J. Yu, and X. Tong (2023). Socialâemotional skills correlate with reading ability among typically developing readers: A meta-analysis. *Education Sciences* 13 (2), 220

5

Appendix



1

Tables

Table 2: Study Groups

Study group	Device(s) children have access to	Encouraged show	Encouraged Device(s)
TV schools: encouraged to watch the TV broadcast of the show	Television	The Nuzo and Namia Show	Television
Control	None	None	None

Table 4: Asset Characteristics

	Treatment	Control
Baseline		
Proportion of households with a smartphone	0.92 (0.27)	0.90 (0.30)
Proportion of households with a TV	0.59 (0.49)	0.56 (0.50)
Asset Score	4.02 (1.66)	3.96 (1.73)
Endline	Treatment	Control
Proportion of households with a smartphone	0.69 (0.46)	0.65 (0.48)
Proportion of households with a TV	0.85 (0.36)	0.82 (0.38)
Asset Score	1.01(0.87)	0.98(0.90)

Table 6: Mean and Standard Deviation of sample characteristics at Endline

Variable	Mean	SD	N
Female	0.50	0.50	3662
Grade 1	0.31	0.46	3612
Grade 2	0.34	0.47	3612
Grade 3	0.35	0.48	3612
TV Hours per week	1.23	0.61	3669
Watches > 30 Minutes of TV	0.87	0.34	3669
Watches Akili Kids	0.86	0.35	3669
HH Size	5.83	2.57	3668
HH has TV	0.84	0.37	3669
HH has smartphone	0.67	0.47	3669
Caregiver has primary education or less	0.04	0.20	3661
HH Speaks English	0.17	0.37	3669
Number of books	16.86	13.58	3669

Table 9: Correlation of gender and educational level of a caregiver with SMS engagement

	Probability of ever responding	Probability of at least one valid response	Number of responses	Number of valid responses
Female	0.043* (0.021)	0.043+ (0.022)	1.420* (0.623)	1.346* (0.607)
Education Level				
Some Primary	0.211+ (0.116)	0.249* (0.116)	4.186 (4.414)	4.350 (4.250)
Primary pass	0.115** (0.043)	0.135** (0.043)	4.285*** (1.015)	4.085*** (0.989)
Secondary pass	0.171*** (0.045)	0.199*** (0.045)	5.783*** (1.098)	5.539*** (1.070)
College/university	0.172*** (0.052)	0.178*** (0.053)	7.424*** (1.410)	7.201*** (1.379)
Postgraduate/Masters	0.367*** (0.048)	0.407*** (0.047)	15.848 (17.008)	15.721 (17.669)
Assets				
TV ownership	0.031 (0.040)	0.017 (0.040)	0.528 (1.130)	0.714 (1.077)
Smartphone Ownership	-0.028 (0.021)	-0.036+ (0.021)	-0.539 (0.622)	-0.522 (0.607)
Phone (non-smart) ownership	0.025 (0.026)	0.025 (0.026)	0.573 (0.747)	0.424 (0.733)
Electricity	0.028 (0.023)	0.030 (0.023)	1.377* (0.662)	1.337* (0.646)
County				
Kakamega	0.026 (0.045)	0.006 (0.045)	1.143 (1.389)	0.800 (1.361)
Kiambu	-0.056 (0.045)	-0.072 (0.045)	-2.712+ (1.388)	-2.801* (1.361)
Kilifi	-0.049 (0.052)	-0.067 (0.052)	-1.327 (1.543)	-1.477 (1.508)
Meru	-0.047 (0.052)	-0.075 (0.053)	-2.859+ (1.542)	-2.880+ (1.515)
Kisumu	-0.041 (0.047)	-0.067 (0.048)	-0.527 (1.457)	-0.638 (1.427)
Num.Obs.	2060	2060	2060	2060
R2	0.022	0.024	0.035	0.034

Note: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Reference county is Kajiado.

Table 12: Ubongo Curriculum for the Nuzo and Namia show

Episode	Learning Objectives
Episode 1	Creating Images/Visualize
Episode 2	Skimming
Episode 3	Questioning
Episode 4	Recalling
Episode 5	Predicting
Episode 6	Connecting
Episode 7	Creating Images/Visualize
Episode 8	Skimming
Episode 9	Questioning
Episode 10	Recalling
Episode 11	Predicting
Episode 12	Connecting
Episode 13	Overview of learning objectives in Nuzo and Namia

2

Figures

Figure 7: Proportion of multi-child and single-child households at Baseline and Endline

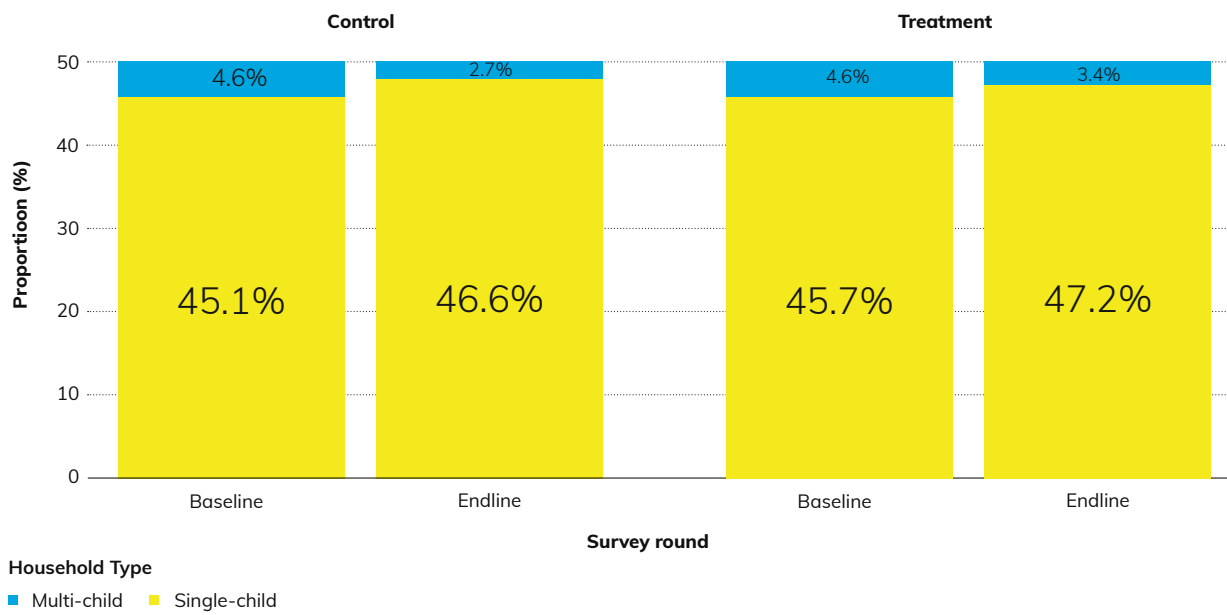
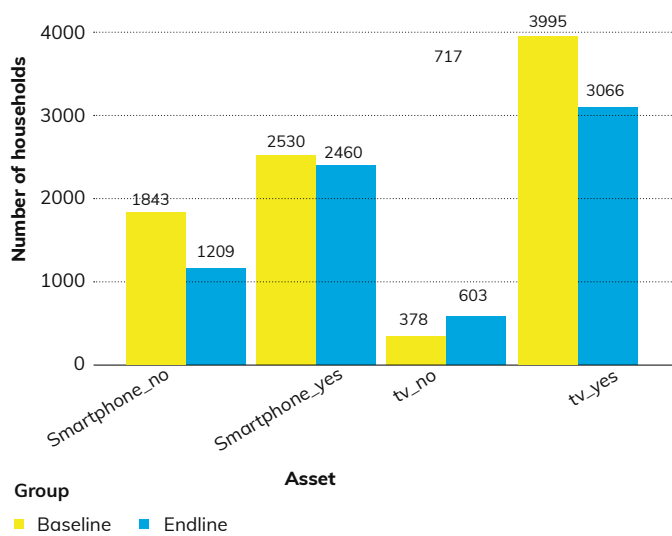


Figure 10: Household Asset Ownership at Baseline and Endline



3

Appendix II: Resources

Ubongo Parent Materials

1. Planner



NUZO

and Namia



RATIBA YA KILA JUMAMOSI		RATIBA YA KILA JUMATANO	
Masaa 	Kinachofanyika kabla ya kutazama "Nuzo and Namia" <hr/> <hr/> <hr/>	Masaa  	Kinachofanyika kabla ya kutazama "Nuzo and Namia" <hr/> <hr/> <hr/>
Masaa  	Kufurahia na kusoma na Nuzo na Namia kwenye Akili Kids 	Masaa  	Kufurahia na kusoma na Nuzo na Namia kwenye Akili Kids 
Masaa 	Kinachofanyika baada ya kutazama "Nuzo and Namia" <hr/> <hr/> <hr/>	Masaa 	Kinachofanyika baada ya kutazama "Nuzo and Namia" <hr/> <hr/> <hr/>



2. Commitment Form



Fomu ya Mzazi kujaza

93% ya wazazi/wanaotoa huduma kwa wazazi/walezi ambao wanawao huona kipindi cha Nuzo and Namia wamesema kwamba kipindi hiki ni bora!



Naweza kusema kipindi cha Nuzo and Namia kimeboresha maisha yangu kwa sababu watoto wangu wanapata elimu bila kulipia masomo ya ziada.

Baba ya watoto watatu walio na miaka 3, 5 na 7.



Mimi _____

na apa kwamba wanangu

watakuwa wakitazama kipindi cha Nuzo and Namia kila Jumamosi saa 10.30 - 11.00 asubuhi katika Akili Kids Channel.



3. TV Instructions

Kipindi cha

NUZO and Namia

by ubongo.

★

**Kinaletwa ndani ya Akili Kids Channel
kila Jumatano saa 6.30–7.00 jioni na
Jumamosi saa 10.30–11.00 asubuhi.**

Usipitwe na Kipindi cha Nuzo and Namia.

Fuata hatua zifuatazo ikiwa usajili wako umeisha, na hauna SMART TV ili wanako wasipitwe na Kipindi cha Nuzo and Namia.

1



Ondoa sanduku la decoda. Sanduku hili linapatikana katikati ya antena na televisheni.

2



Wekelea antenna uliochukua kwenye sanduku kwa antenna ya televisheni

3



Zima televisheni na ungoje.

4



Fungua televisheni na tafuta chaneli za analogue.

5



Chagua Akili Kids.

Tulia na usome, utatue na ufurahie Nuzo and Namia na marafiki!



4. Watching Tracker



Ratiba ya

NUZO and Namia

★

Inafaa utazame Nuzo and Namia angalau mara moja kila Jumamosi au Jumatano.

Weka alama kwenye Jumamosi au Jumatano ambazo mtoto/watoto wako wametazama "Nuzo and Namia".

Jumamosi

Juni	Tarehe 24				
Julai	Tarehe 1	Tarehe 8	Tarehe 15	Tarehe 22	Tarehe 29
Agosti	Tarehe 5	Tarehe 12	Tarehe 19	Tarehe 26	
Septemba	Tarehe 2	Tarehe 9	Tarehe 16	Tarehe 23	Tarehe 30
Oktoba	Tarehe 7	Tarehe 14	Tarehe 21	Tarehe 28	
Novemba	Tarehe 4	Tarehe 11	Tarehe 18	Tarehe 25	

Jumatano

Juni	Tarehe 21	Tarehe 28			
Julai	Tarehe 5	Tarehe 12	Tarehe 19	Tarehe 26	
Agosti	Tarehe 2	Tarehe 9	Tarehe 16	Tarehe 23	Tarehe 30
Septemba	Tarehe 6	Tarehe 13	Tarehe 20	Tarehe 27	
Oktoba	Tarehe 4	Tarehe 11	Tarehe 18	Tarehe 25	
Novemba	Tarehe 1	Tarehe 8	Tarehe 15	Tarehe 22	Tarehe 29

5. Parent Poster

**NUZO
and
Namia**

Time to have fun and learn

**6.30pm – 7.00pm every Wednesday
and 10.30am – 11.00am every Saturday
on the Akili Kids Channel**

**Help Namia find her way to Nuzo
in time for the next episode !**

